

Carbon impacts made visible

Despite disagreement between governments about tackling climate change, initiatives are bubbling up from below. With help from researchers and the markets, citizens can be made more aware of how to help reduce carbon dioxide emissions.

Europeans are well aware of the need to improve their responses to extreme weather. After the floods, heatwaves and droughts of the past five years, many governments and regions are developing their capacity to issue forecasts and heighten the alarmingly poor awareness of citizens both of their vulnerability and of what to do in the event of catastrophes. They are also spending big money on protecting natural systems to mitigate the effects of peaks in rainfall and spells of high temperature, both of which are on the increase.

The case of the River Tisza in eastern Europe, which experienced an unprecedented series of floods in 1998–2001, highlights how far some regions have come. A high level upstream in 2000 gave Hungarians two weeks' notice of a potentially disastrous rise downstream, requiring more than a million sandbags to be laid every day to stop the countryside flooding. As reported at a meeting hosted by the Environmental Protection Agency of Ireland in Dublin last week to assess the effects and communication of climate change, a €500-million (US\$600-million) development plan and awareness campaign is in place, including the construction of water-retention basins. Similar developments in Poland, Britain and elsewhere are complemented at a European level by large-scale modelling by the Joint Research Centre, intended to provide longer-term warnings of threats.

So far, so impressive. But few of the national initiatives have been placed within a framework of decadal climate trends. Perhaps this is not surprising given that regional climate models are still full of uncertainties. But this only underlines the key challenge: how to convince citizens to do their bit to moderate the unimaginably enormous scale of mankind's carbon dioxide emissions.

Whether or not the Kyoto Protocol is ratified, at least two bottom-up opportunities must be grasped. One lies in the Kyoto-inspired, but not necessarily Kyoto-dependent, market for emissions trading (see *Phil. Trans. R. Soc. A*, August 2002 special issue). A combination of a voluntary domestic market in the United States, European Union initiatives, and national and state schemes worldwide has yielded a rapidly growing but still pitifully small market, currently trading some 70 million tonnes of carbon dioxide per year. The extent to which such schemes can alleviate emissions without Kyoto-like treaties is not clear. But they can lead to a change in habits as awareness of the costs of carbon emissions permeates through the companies, organizations, regions and investors involved. A key step is to incorporate air transport, land use and carbon sequestration into trading frameworks. This should reduce the inequities, especially if they allow the developing world to profit more sustainably from its natural resources.

Advocates of responsible behaviour must seize every opportunity to get their message across, such as the forthcoming ice-age blockbuster *The Day After Tomorrow*, due for release worldwide on 28 May. The film's website shows how to offset the carbon impact of your climatologically reckless lifestyle. Climatologists will criticize the faulty science on which the film is based, but it will allow them to raise citizens' awareness of the state of climate and ocean science and, moreover, to heighten carbon consciousness. Climate researchers should contact their local media, who will seize the chance to trade on a disaster movie while tapping into the public fascination with science. If, as a result, the graph of carbon dioxide's atmospheric increase can be seen as our era's ticking clock, so much the better. ■

Ethics of therapeutic cloning

A moment of triumph for South Korean science appears to have been marred by doubts about lab practice.

Therapeutic cloning is one of the most divisive topics in modern biology. To some, it promises a future in which damaged and diseased tissues and organs will be replaced without worrying about immune rejection. To others, the idea of creating a human embryo and culturing it for several days to obtain stem cells that would be needed to grow such grafts is morally reprehensible. The two sides have been battling it out in legislatures across the world over the past few years.

The last thing those engaged in therapeutic cloning research need, therefore, is further ethical controversy. Yet a storm seems about to break over the field's most prominent paper to date — the report in February that a group in South Korea had derived a line of embryonic stem cells from a cloned human embryo (W. S. Hwang *et al. Science* **303**, 1669–1674; 2004). Questions are being raised about how the researchers managed to recruit so many women prepared to donate their eggs for the project. One such question is why a PhD student, who was a co-author on the paper, initially told *Nature* she was an egg donor but later changed her story (see pages 3 and 12).

In the context of South Korean society, it's easy to see how students involved in such a project might, with the best of intentions, want to

donate their eggs. Korea is an intensely patriotic country, in which the desire to help others is deeply ingrained. The prospect of helping sick patients, and demonstrating that Korea is capable of world-leading research, would be a powerful motivating factor. In such circumstances, say bioethicists, procedures should be in place to ensure that the donors are all volunteers with no direct connection to the research. The principal investigators must now demonstrate that such safeguards were in place, and that they were rigorously applied.

Questions should also be asked of the local Institutional Review Board (IRB) that gave ethical approval for the research project. IRBs are supposed "to assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of humans participating as subjects", according to the US Food and Drug Administration. So far, the IRB that approved the Korean cloning project has been less than forthcoming about its work.

If the air is not cleared quickly, the consequences for Korean science — and for research into therapeutic cloning internationally — could be severe. It will be a tragedy if one of the greatest scientific stories of the year ends up being remembered, in South Korea especially, as one that lost the trust of the people. ■





Robodocs
NASA looks to
automation to keep
Hubble out of trouble
p4



Rock show
Fossil exhibition
earns mixed response
from museums
p5



Up in the air
Microwaves offer
clues to Earth's slow
cooking time
p7



Isaac who?
Web fame test brings
physicists down to
Earth with a bump
p8

Korea's stem-cell stars dogged by suspicion of ethical breach

David Cyranoski, Seoul

When, in February, a South Korean team announced that it had derived stem cells from a cloned human embryo, its achievement was heralded as an important step on the road to 'therapeutic cloning'. But the research is now clouded by nagging questions about the source of the key resource for the experiment: human egg cells.

Korean citizens'-rights activists and bioethicists are pressing the team, led by Woo Suk Hwang and Shin Yong Moon of Seoul National University, to prove that the recruitment of women volunteers followed ethical guidelines. *Nature's* enquiries have also revealed troubling inconsistencies — in particular over whether the donors included junior members of the research team.

The brewing controversy could undermine the domestic public and political support on which Hwang and Moon's progress has depended (see News Feature, page 12). Any suggestion of ethical irregularities in therapeutic-cloning research could also have international repercussions, providing ammunition for activists who are opposed to the technology on moral grounds.

Therapeutic cloning involves creating an embryo by transferring the nucleus from a patient's cell into a human egg cell stripped of its own nucleus. After being grown in culture for a few days, this clone can yield embryonic stem cells, which can develop into any of the body's tissues. Because these would be derived from the patient's own cells, there should be no problem with immune rejection in using grafts derived from them to repair diseased or damaged tissues.

But cloning is very inefficient. To derive a single line of embryonic stem cells, the Korean team used 242 eggs obtained from 16 volunteers (W. S. Hwang *et al. Science* **303**, 1669–1674; 2004). Each woman was given hormone injections to force her ovaries to superovulate, producing 12–20 eggs per menstrual cycle instead of one.

Other researchers were surprised that so many women were prepared to undergo this procedure for a research project. Side effects of the treatment can range from general discomfort and emotional stress to clotting of



Woo Suk Hwang faces increasing scrutiny over how he recruited egg donors for his research.

the veins or stroke. "It's a painful procedure and there is risk involved," says Jose Cibelli, a co-author on the paper who studies cloning at Michigan State University in East Lansing. "It would never fly in the United States."

Hwang says that the donors were motivated by a desire to push forward a promising field of medicine. "Many women are sympathetic with our research," he says. Supplementary material published online with the paper says that the volunteers were not paid, and explains that they filled in informed-consent forms detailing how the eggs would be used.

Egg donations

The donors were anonymous, but one PhD student in the team, Ja Min Koo, initially told *Nature* that the donors included herself and another woman in the lab. She subsequently called back and said that she had not donated eggs, blaming her poor English for a misunderstanding. But in the initial interview, she named the hospital where her donation was carried out, and explained that she had been happy to donate eggs because she already has two children.

Art Caplan, who heads the Center for Bioethics at the University of Pennsylvania in Philadelphia, argues that it would be bad practice if egg donors for such a project

included students or junior employees on the research team because "it could certainly look like coercion was involved".

The information posted with the paper also states: "Neither donors nor their family, relatives or associates may benefit from this research." Koo, who was a co-author on the paper, arguably did stand to gain professionally from its publication.

Hwang denies that Koo was among the donors. But he declined *Nature's* requests for further documentary evidence of the procedures for recruiting the egg donors and obtaining their consent. Attempts to get more information from the Institutional Review Board at Hanyang University Hospital in Seoul, which provided ethical approval, were similarly rebuffed. Its chair, university obstetrician Moon-il Park, cancelled an arranged phone interview.

Within Korea, concern is growing about the lack of transparency surrounding the procedures for obtaining the donated eggs. "I'm doubtful women would give their eggs so easily," wrote Pil Pyul Lee, a science historian at the Korea National Open University in Seoul, in the 23 February issue of the *Professors Times*, a nationwide newspaper in which academics express their views on topical issues.

Lee's article also questioned the inclusion as a co-author on the paper of Ky Young Park, a plant molecular biologist formerly at Sunchon National University who is now a science and technology adviser to South Korean President Roh Moo-hyun. Park says that she has played an important role in Hwang's research over the years by advising him on public attitudes to his work with transgenic livestock, but she told *Nature* that she had no specific involvement with the therapeutic-cloning paper.

One of South Korea's leading citizens'-rights groups, People's Solidarity for Participatory Democracy, now says that it will look into the ethical issues surrounding the cloning paper. "We plan to pressure the government to force them to produce the documents that are required," says Jae-kak Han, who heads the group's scientific division. The Korean Bioethics Association is also pressuring the National Human Rights Commission to look into the matter. ■

BioShield defence programme set to fund anthrax vaccine

Erika Check, Washington

A US government programme to bolster public defences against bioterrorism will soon give out its first grants — despite complaints from the biotechnology industry that the programme is not working as intended.

An official at the Department of Health and Human Services says it is reviewing proposals for a next-generation anthrax vaccine and expects to fund one of them this year under the BioShield programme. Congress allocated \$890 million to the programme in 2004, but biotechnology companies say the money has not materialized.

Project BioShield was first proposed by President George Bush in his State of the Union Address in January 2003. He said the programme would spend \$5.6 billion over ten years on developing products, such as vaccines, that could be used in response to bioterrorist attacks. Bush said the money would give companies an incentive to test and manufacture products that might not otherwise have a commercial market.

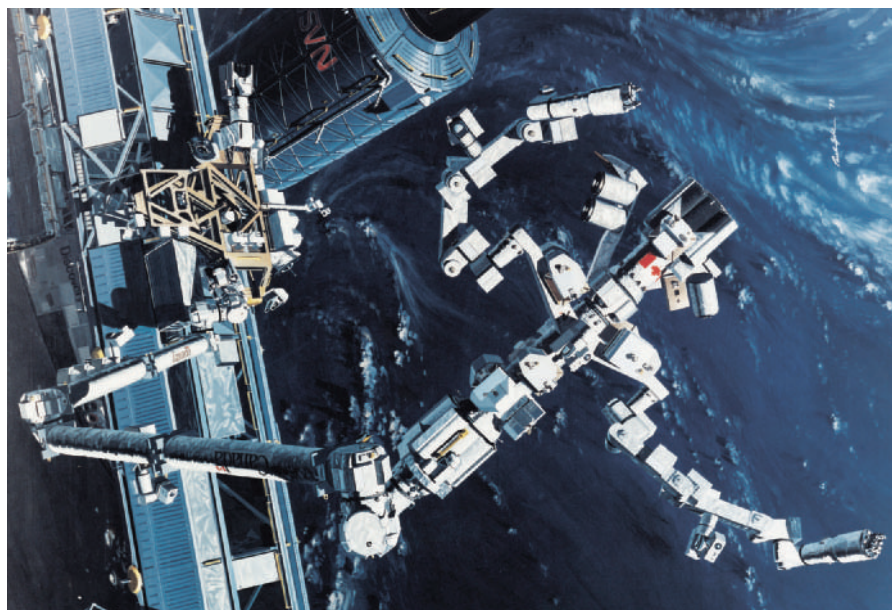
But even though Congress has given the health department funding for BioShield in 2004, it has not yet passed legislation setting out detailed rules for how the money should be spent, leaving many questions unanswered about the programme's future.

Companies have complained about the delay in passing the legislation. They say that uncertainty over the programme is stopping them from conducting late-stage trials of potential therapies. For instance, Human Genome Sciences of Rockville, Maryland, says that it has halted development of a monoclonal antibody that could be used against anthrax.

"We have delivered what we thought the government was looking for, but we are at a point now where we would require a considerably larger investment and we are not able to do that without a government commitment," says Jerry Parrott, a spokesman for the company.

But officials at the health department say the rules for implementing BioShield are already clear. They say the 2004 funding legislation laid out some guidance for the programme, and that department officials already have the experience needed to grant BioShield contracts.

"This is important and the money is there, so we feel we would be remiss if we did not use it as Congress intended," says one department official. ■



Automatic for the people: machines could take the place of humans to help Hubble.

NASA opens its arms to robot options for saving telescope

Tony Reichhardt, Washington

A plan to use robots instead of astronauts to rescue the Hubble Space Telescope moved closer to approval last week, when NASA announced that it would solicit industry proposals for part of a salvage mission.

Scientists who were initially sceptical of relying on robots for the job say it now looks feasible — provided the agency comes up with several hundred millions of dollars to pay for it.

Engineers from seven NASA centres will gather at the Goddard Space Flight Center in Maryland on 13–14 May to review plans, recently developed there, to service the telescope robotically. Hubble's batteries are expected to fail at some point in 2008 or 2009.

If the review is favourable, NASA will request bids on 1 June for robotically de-orbiting the telescope to prevent it crashing to Earth. And, says Hubble programme manager Preston Burch, the agency is now "very optimistic" about attaching a robot to upgrade Hubble's instruments and extend its lifetime (see *Nature* 428, 353; 2004).

NASA officials seem to prefer the Special Purpose Dexterous Manipulator (Dextre), a two-armed robot built by MD Robotics of Brampton, Ontario, which they plan to install on the International Space Station in 2007. Other contenders for the job are not technically mature or have never flown in space, says Burch.

Dextre is already fully built and was derived from proven hardware: the space shuttle's robot arm, which has been used successfully for more than 20 years. MD Robotics

has loaned Goddard the use of test equipment, and initial simulations of Hubble repair tasks have been very encouraging, says Burch.

Finance remains an issue, however. Even the deorbit-only mission — which involves attaching a propulsion module to Hubble and steering it into the ocean — has an estimated cost of at least US\$500 million, including a launch vehicle. Adding Dextre or another robot to service the telescope would require a bigger launch vehicle, which would raise the price, says Burch. But robotic servicing would still be cheaper than sending space shuttle astronauts to do the job.

Steven Beckwith, director of the Space Telescope Science Institute in Baltimore, worries that NASA might settle for "half a loaf" and only do the bare-bones deorbit or minimal repairs, as he estimates that a full servicing mission will cost "a large fraction of a billion dollars". But although he and other astronomers originally argued that astronauts were necessary for the full servicing job, they now accept that the robotic repair is technically feasible. According to Burch, a briefing of the Hubble Space Telescope Users Committee late last month convinced members that the technology could work.

Beckwith still thinks astronaut servicing has the "highest probability of succeeding", however. A National Academies panel formed to look at options for extending Hubble's life will consider both human and robotic servicing. The panel, chaired by Louis Lanzerotti, an astrophysicist and consultant with Bell Laboratories, will have its first meeting next month and will report in November. ■

Feathered fossils cause a flap in museums

Rex Dalton, San Diego

The mysterious background of Chinese dinosaur fossils in an exhibition is posing awkward questions for North American museums, which want to ensure that only properly obtained specimens are displayed.

Experts and Chinese officials are worried that some of the 34 fossils in the exhibit — currently on show at the San Diego Natural History Museum — may have been smuggled out of China and sold in the underground trade. The Royal Ontario Museum in Toronto is investigating the allegations before deciding whether to display the feathered specimens next year.

Stephen and Sylvia Czerkas, the Utah-based married couple who compiled the exhibit, say they have permission to borrow the fossils from China, and will return them when the exhibit has finished touring. "We did not purchase them," says Sylvia Czerkas.

But Ji Qiang, head palaeontologist at the Institute of Geology in Beijing, China, thinks that about a dozen of the specimens may be in the United States improperly. When he first saw the fossils three years ago at the Czerkases' Dinosaur Museum in Blanding, Utah, Ji says he identified three of them as being of "great scientific value".

After discussions with the Czerkases, Ji says, he assigned specimen numbers to the three most desired fossils for his geological institute.

Ji says he is in no doubt that the specimens were allowed to leave China illegally. "The problem is, how do I do research on them?" he asks. "And how can I get the specimens?"

Sylvia Czerkas says that all of the specimens are to be returned to China in 2007 under a 2001 loan agreement reached with officials of Liaoning province in north-east China, where fabulous fossils from 120 million years ago have been found.



Early birds? Feathered dinosaurs such as *Caudipteryx* provide clues to how flight evolved.

When informed that Ji had described the specimens as "illegal", Czerkas responded: "That is not true."

Stephen Czerkas is an artist who is self-taught in palaeontology, and Sylvia serves as museum curator. Five years ago, the couple were involved in an international controversy after their museum bought a fossil, called *Archaeoraptor*, for \$80,000 at a fossil show in Tucson, Arizona.

The fossil appeared in *National Geographic* magazine after failed attempts to publish it in both *Nature* and *Science*. But it was subsequently found to be a forged composite from different species (see *Nature* 410, 539; 2001), put together in China to resemble a 'missing link' between dinosaurs and birds. It was returned there in 2000, and

segments later generated scientific papers by Chinese authors and by Stephen Czerkas.

The Czerkases have been offering their latest exhibit, called "Feathered Dinosaurs and the Origin of Flight", to interested museums for two years. The museums pay a set fee for the display. Dinosaur exhibits are in demand because they bring in the crowds, but some museums say that they have declined to take the exhibit.

At the Natural History Museum of Los Angeles County, which didn't consider the Czerkas exhibit, palaeontology curator Luis Chiappe says that specimens that come from the underground trade should not be displayed, even if they are to be repatriated. Exhibiting them promotes commerce in illegal specimens, he says, and encourages the removal of more palaeontological treasures.

In its first showing outside their own museum, the exhibit opened a seven-month run in February at the San Diego museum. Palaeontologist Mick Hager, the museum's executive director, says he believes that the specimens are important and should be displayed. He adds that he accepts the Czerkases' assurances that the specimens are legally loaned and will be returned to China.

Two Chinese officials from Liaoning attended the opening, but the Czerkases refused requests to make them available for interview. Ji was billed as due to attend, but chose not to do so.

Mark Engstrom, the Royal Ontario's vice-president for collections, says it has opened an investigation into the fossils' recent history. The museum has agreed to bring the fossils to Toronto in the spring of 2005, but Engstrom says his facility will not fully commit until more is known about them. "It is up to the museum community to police itself," he says. "We have to determine if the material can be legally imported to Canada." ■

Top job at NSF on hold until after US elections

Geoff Brumfiel, Washington

Prospects are fading that a director for the US National Science Foundation (NSF) will be named any time this year, officials familiar with the search say.

The top slot at the research agency has been filled on an interim basis by Arden Bement, head of the National Institute of Standards and Technology, since Rita Colwell vacated it on 21 February (see *Nature* 427, 665; 2004). On 26 February, John Marburger, head of the White House Office of Science and Technology Policy, told a congressional hearing that he was "very optimistic" that a replacement

could be found in a matter of months.

But officials close to the search process think that the frenzy of election year will make this unlikely. "The longer it takes them to announce a candidate, the less likely it is they can get him or her confirmed," says Joel Widder, a lobbyist with Lewis-Burke Associates in Washington and former head of legislative affairs at the NSF.

The gap leaves plenty of time for science lobbyists to speculate on who Colwell's eventual replacement will be. One name said to be in the frame, if President Bush is re-elected in November, is Raymond Bowen. A mechanical engineer at Texas A&M

University in College Station, Bowen has been a member of the National Science Board — the presidentially appointed body that oversees the NSF — since 2002.

Bowen served as president of Texas A&M from 1994 to 2002. He worked at the NSF as head of its mechanical-engineering division from 1982 to 1983, and as deputy chief of its engineering directorate from 1990 to 1991.

Bowen says he has not been officially approached for the job, and refuses to say whether, if offered, he would take it. "I am deeply involved in preparing an exam for a class that needs to be challenged in ordinary differential equations," he told *Nature*. ■

Tax change curtails UK university spin-offs

Jim Giles, London

The creation of spin-off companies at British universities has fallen dramatically over the past six months because of a change in the tax laws, *Nature* has established.

The country's top universities typically each launch five to ten new companies every year. But the four highest-ranked research institutions — University College London, Imperial College London and the universities of Oxford and Cambridge — have created just six spin-offs between them since last August, when the new rules came into force.

The spin-offs have dried up, university officials say, because the tax laws now require directors to pay hefty taxes on the estimated value of shares they hold in a company within a year of its foundation. The rule was meant to close a tax loophole used by wealthy business executives — but it has had the unintended effect of making scientists think twice about starting a business venture.

"The rules have had a devastating effect," says Tom Hockaday, executive director of Isis Innovation, the University of Oxford's technology-transfer company. "They're killing what they say they want to create."

Sue Sundstrom, director of life sciences at the University of Southampton's Centre for Enterprise and Innovation, says the tax bill is "completely insane" given that government ministers are always calling on universities to do more to create wealth.

The drought might ease soon, because



All quiet: attempts to close a tax loophole have slowed activity at places such as Oxford Science Park.

university technology-transfer officers and tax officials have agreed on a fresh interpretation of the law. But university officials caution that the revised interpretation will pose further problems for academics, and won't guarantee a full revival of spin-off activity.

In the revised arrangement, researchers can defer paying taxes by taking 'convertible preference' shares, which are subject to tax only at the time they are sold.

Technology-transfer officers say the change may encourage some academics to

start launching companies again, but that it will expose them to even higher tax rates — of up to 48% — when they sell their shares.

A spokesman for the Inland Revenue, which helped draft the tax laws and negotiated the new arrangements with universities, said he recognized that there "had been a period of uncertainty" but that the revision should help universities meet commercial goals without creating tax problems for academics. He declined to comment on whether the higher tax burden would eventually be removed. ■

AIDS drug price hike prompts calls for intervention

Erika Check, Washington

The National Institutes of Health (NIH) has been asked to step in and halt the spiralling cost of AIDS medications in the United States.

Under a 1980 technology-transfer law, the biomedical research agency can arrange for the generic production of pharmaceuticals whose invention was based on NIH-funded research, if the existing manufacturer will not provide the treatment on "reasonable terms".

In January, Essential Inventions, a non-profit group based in Washington DC, asked the NIH to invoke these 'march-in' rights with regard to the AIDS drug Norvir, made by Abbott Laboratories of Abbott Park, Illinois. Last December, Abbott raised the US price of Norvir treatment from \$1.71 to \$8.57 per day.

The NIH will discuss the request at a public hearing at its headquarters in Bethesda, Maryland, on 25 May. Abbott says the price rise is necessary because the drug is used in low doses, resulting in lower sales.

The request's supporters include Sherrod Brown (Democrat, Ohio), the senior minority member of the House of Representatives subcommittee, which oversees the NIH. "When a company raises the price of an essential medicine by 400%, it gets anybody's attention," Brown says. Taxpayers pay for biomedical research, he says, "but we never invoke the mechanism that would bring prices down for the public".

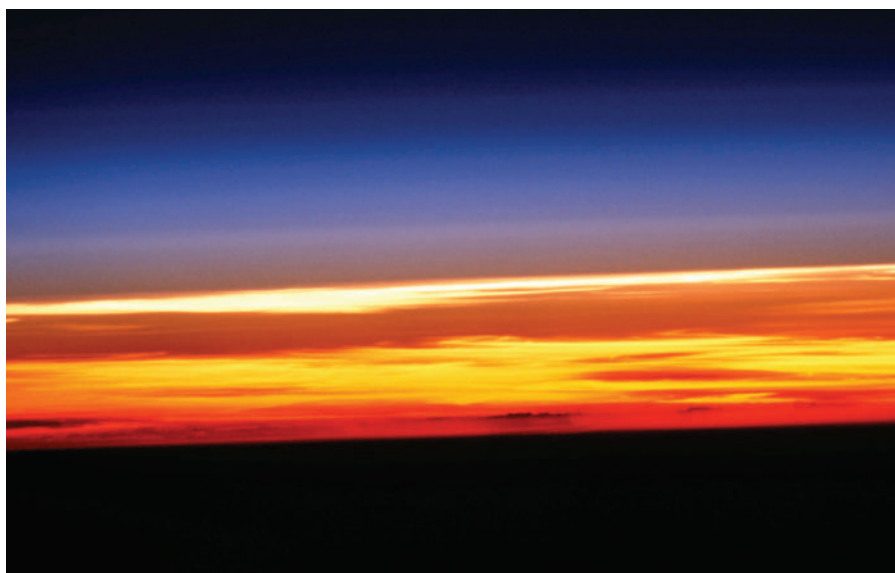
The price increase angered AIDS activists because Norvir boosts other AIDS drugs called protease inhibitors — so all patients taking these are affected. The issue also coincides with a national argument over drug pricing in the run-up to November's elections.

That environment could make it hard for the NIH and its parent agency, the health department, to turn down the request. Health secretary Tommy Thompson "would have to say yes, it is reasonable for an invention to be ten times more expensive here than in other countries, even if US taxpayers funded it," says James Love, president of Essential Inventions.

Abbott says the price increase will not affect uninsured patients, who get the drug free, or others who get it under government programmes.

The company's supporters say that the Bayh-Dole Act, which contains the march-in clause, was not meant to control drug prices. Joseph Allen, president of the National Technology Transfer Center in Wheeling, West Virginia, and a former Senate staff member who worked on the act, says the phrase "reasonable terms" was not supposed to embrace price increases.

Love agrees that the question of what constitutes "reasonable terms" will be crucial to the outcome of the request. But he argues that Abbott is trying to quash competition by charging more to patients who use Norvir to boost other protease inhibitors, rather than using Abbott's own treatments, which incorporate the drug. "This is designed to prevent people from buying non-Abbott protease inhibitors," he says. "If this doesn't trigger the march-in, nothing does." ■



Sky high: cooling in the stratosphere influences temperature trends in the layer of atmosphere below.

Global warming anomaly may succumb to microwave study

Quirin Schiermeier

For years, climate researchers have struggled with an apparent discrepancy in the data on global warming: temperatures in the lower atmosphere have been rising far slower than models predict, given how fast the Earth's surface is heating.

The discrepancy has been central to the arguments of sceptics about global warming. But according to a study in this issue of *Nature* (see page 55) it can be explained by interactions between the troposphere — the first 11 km of the atmosphere — and the stratosphere above it.

In the study, a team from the University of Washington at Seattle and the Air Resources Laboratory of the National Oceanic and Atmospheric Administration (NOAA), based in Maryland, analysed microwave emissions from the atmosphere. The emissions were recorded between 1979 and 2001 by NOAA's polar orbiting satellites. The data can be used to deduce temperatures in different layers of the atmosphere. And the study finds that stratospheric cooling, a known effect of greenhouse gases, appears to account for discrepancies between temperature trends on the ground and in the troposphere.

The team, led by Qiang Fu, an atmospheric researcher at the University of Washington, subtracted the impact of such cooling from data on the stratosphere and performed a statistical analysis, which found temperature trends consistent with observed warming on the surface and the predictions of climate models.

The finding is "a stunningly elegant and accurate method of clarifying global trends", says Kevin Trenberth, head of the climate analysis section at the National Center for Atmospheric Research in Boulder, Colorado.

But it does not impress John Christy, director of the Earth System Science Center at the University of Alabama in Huntsville, whose work established the inconsistency between temperature trends on the surface and in the troposphere (*Science* **247**, 1558–1662; 1990). "You cannot eliminate the stratospheric influence with statistical tools alone," he says. "If you want to know precisely what happens you need physical measurements." He says that Fu has overcorrected for the impact of the stratosphere in his analysis.

Other climate scientists welcomed the new findings. "This is the answer — I wish we had recognized it ourselves," says John Wallace, an atmospheric researcher also at the University of Washington, who chaired a 2000 survey on reconciling global warming discrepancies for the US National Academies.

The study should be noted by policy-makers who justify lack of action on global warming by citing scientific uncertainty, says Wallace. But he is not optimistic about how many minds will be changed. "Single scientific discoveries have little impact in the political arena," he says. NOAA officials declined to comment on the political implications of the study.

The research comes just after a report from the Pew Center on Global Climate Change predicted that global warming could shrink the US economy.

But neither economic nor scientific analyses are likely to affect US climate change policy, says Henry Jacoby, director of the Massachusetts Institute of Technology's Joint Program on the Science and Policy of Global Change. "After the Kyoto fiasco, the US administration began to ask for advice from all sides," he says. "But unfortunately it has never taken the advice it received." ■

Fatal fruit bat virus sparks epidemics in southern Asia

Declan Butler

Television cameras may be few and far between in rural areas of Bangladesh, south Asia's poorest nation. But killer human viruses are recurrent there, and are quietly wreaking havoc.

An outbreak of the emerging Nipah virus in the Faridpur district of the country earlier this month infected 30 people and killed 18. That epidemic is one of several that have hit Bangladesh, Malaysia and Singapore since the virus was first discovered in 1998.

The Nipah virus and the related Australian Hendra virus form the *Henipavirus* genus within the paramyxovirus family. They cause high mortality — two in every five infected people die — but pose less of a global risk than more notorious viruses such as SARS (severe acute respiratory syndrome). That is because infection requires close contact with the animal host, the *Pteropus* fruit bat, and does not seem to pass from human to human.

The virus was discovered in Malaysia, during a 1998 outbreak in the village of Nipah that killed 105 people. The transmission route of the virus in Bangladesh seems different from outbreaks elsewhere, however, says Marie-Claude Georges-Courbot, an expert on the disease at the Pasteur Institute in Lyon.

The exact mechanism of transmission is poorly understood, says Georges-Courbot. The Malaysian epidemic arose in humans following an outbreak in pigs that had come into contact with bat urine. But the outbreaks in Bangladesh — one in February killed 17 people — appear to have been caused by children who had direct contact with bat-contaminated fruit, she says.

The symptoms of the disease in Bangladesh are also different, being largely neurological; the Malaysian outbreak caused pulmonary complications.

Georges-Courbot is part of a team of French and Malaysian researchers who earlier this year reported that golden hamsters could be protected from Nipah by vaccinia viruses. The team used vaccinia to express two proteins that the Nipah virus uses to enter human cells. The research also showed that serum from vaccinated hamsters prevented onset of the disease in others. This suggests the possibility of an immunotherapeutic approach to treating Nipah, and perhaps its Hendra cousin. ■

Research ships firing airguns provoke Mexican stand-off

San Diego For the second time this year, US seismic studies in Mexico's waters have been cancelled because of concerns about harm to marine mammals.

Permission for the research ship *Roger Revelle* to use acoustic devices to examine geological rifts off Mexico's western coast was withdrawn on 15 April by the Secretariat of Environment and Natural Resources, an environmental agency of the Mexican government. The cruise, which began on 23 April, will be cut by about a quarter because of the inability to use airguns to fire bursts of compressed air in the sea. The bursts generate sound waves, which are picked up by the ship after reflecting off the sea floor and the rock beneath it.

In February, Mexican officials blocked a similar project using the *Maurice Ewing* off Mexico's eastern coast (see *Nature* **428**, 681; 2004). The 3,500-tonne *Roger Revelle*, which is operated by the Scripps Institution of Oceanography in La Jolla, California, and uses much smaller airguns than the *Maurice Ewing*, recently completed a cruise in the Gulf of California. No marine mammals

were known to have been harmed then, says chief scientist Peter Lonsdale.

Only legendary physicists pass Google's fame test

Washington Heavyweight physicists may think they are public figures, but only giants such as Newton have achieved true fame, according to an analysis using the Google search engine.

Physicist James Bagrow and colleagues



What have you done for me lately? Isaac Newton is only a third as famous as Janet Jackson.

from Clarkson University in Potsdam, New York, compared the number of Google hits for a researcher's name and field with the number of papers that the scientist has published. The correlation was linear. Other groups of renowned figures, such as First World War fighter pilots, show an exponential link between fame and achievement, measured by the number of enemy planes downed.

Only a few physicists, such as Newton, show the same exponential relationship. But even he generates only a third as many hits as the pop star Janet Jackson.

♦ <http://arxiv.org/abs/cond-mat/0404515>

Darwin returns to Italy's evolving curriculum

Rome Italian education minister Letizia Moratti has reversed her earlier decision to remove evolution from the school curriculum of children aged 11 to 14. She also promised that evolutionary theory will be reintroduced for schoolchildren of all ages.

Moratti said in March that biology lessons need not feature evolution, but added that creationism should be a part of voluntary religious-studies classes (see *Nature* **428**, 595; 2004). After being criticized for the move, Moratti set up a

committee on 28 April, headed by the Nobel-prizewinning neurologist Rita Levi Montalcini, to advise about what to teach.

Sidelining Darwin was part of an overhaul of the school curriculum designed to make it less academic and more applicable to daily life.

Popularizer Greenfield is blackballed by peers

London Senior members of the Royal Society have decided not to admit celebrity neuroscientist Susan Greenfield into their elite club.

The fact that Greenfield was being considered for membership of the society caused a stir when it was revealed in February (see *Nature* 427, 578; 2004).

Greenfield is a professor of pharmacology and senior research fellow at the University of Oxford, but is better known for popularizing science on television and in newspapers. Some fellows threatened to resign if she was admitted.

The society declined to go into detail about the nomination process. A spokesman said last week that Greenfield's name had been considered by committees within the society but had not been placed on the list of candidates, which was circulated to existing fellows on 23 April.

Map of swirls and sea monsters is spot on

London Unlikely as it may seem, the swirls drawn next to monsters and galleons on this map of the northeast Atlantic may not be just artistic licence, say remote-sensing experts. The swirls off the east coast of Iceland match thermal images captured by Earth-observation satellites of the same area. The eddies are created where the Gulf Stream meets cold Arctic waters.

The map was drawn by Olaus Magnus, a Swedish priest, in 1539. Ancient mariners would have noticed these large-scale eddies, which pull on shipping vessels and have greener waters. Tom Rossby, an oceanographer at the University of Rhode Island in Kingston, first



noted the similarities between the map and satellite images. He says that the swirls are too deliberate to be artistic expression.

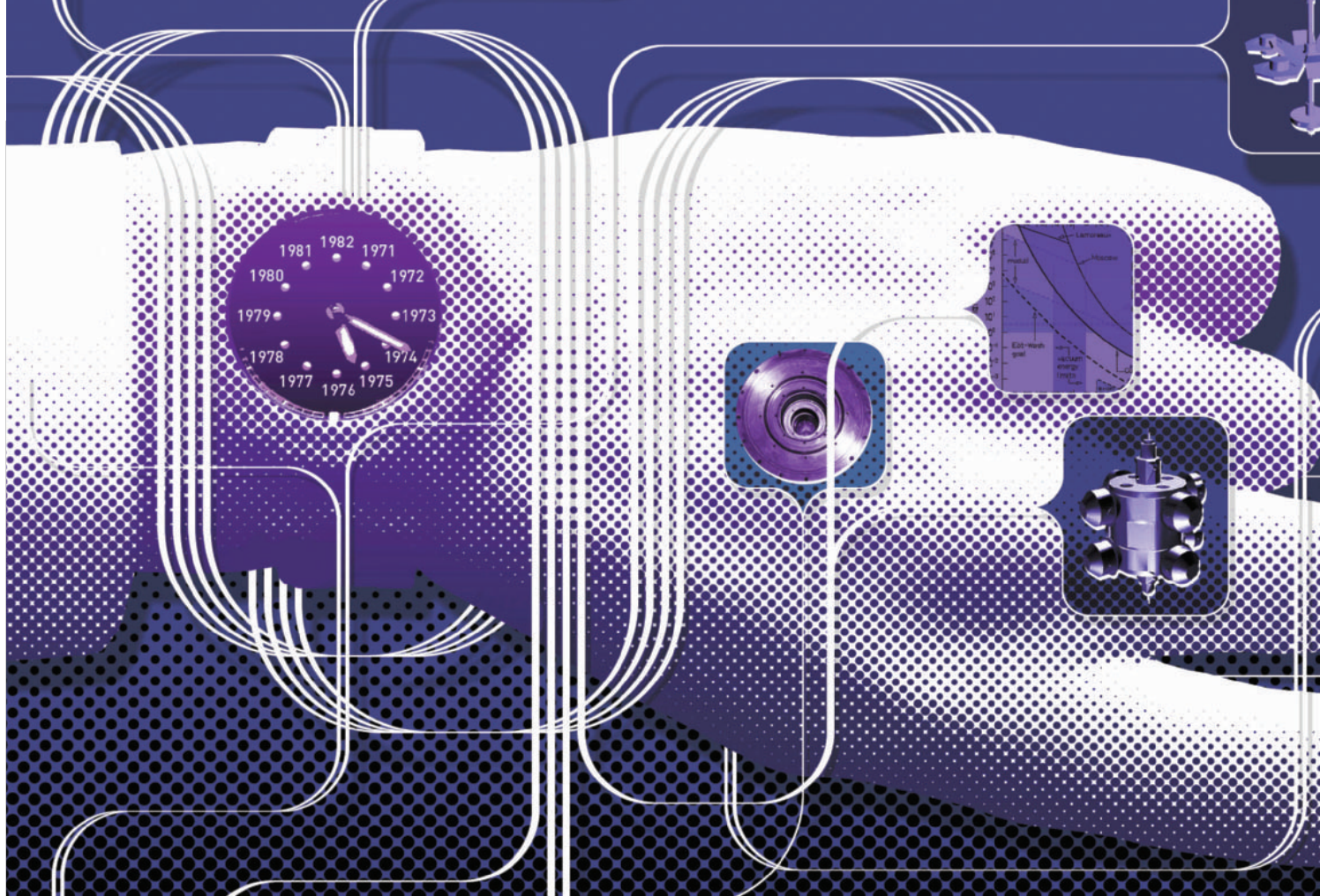
Reinstated cancer chief is sacked again

Munich Having been fired and then apparently reinstated as scientific director of Italy's National Cancer Institute in Genoa, Lucio Luzzatto last week reluctantly left again.

Luzzatto was recruited from the Memorial Sloan-Kettering Cancer Center in New York four years ago. He was sacked on 1 April by the institute's administrative director, Maurizio Mauri, who objected to Luzzatto's continuing collaboration with

Sloan-Kettering scientists. Less than a week later, Mauri agreed to rescind the dismissal, accepting that his grounds had been inappropriate (see *Nature* 428, 683; 2004).

But on 24 April, Mauri said the 'reinstatement' was conditional on Luzzatto offering to resign immediately. Luzzatto, who remains sacked, says he will fight the decision. In an open letter to the Italian health minister Girolamo Sirchia, published in the *La Repubblica* newspaper on 29 April, Luzzatto says he has support from the scientific community at home and abroad.



The waiting game

A few physicists have spent decades searching for the rarest events in the Universe — and seen nothing. But their enthusiasm for the hunt is undimmed. Geoff Brumfiel asks what keeps them going.

Blas Cabrera still remembers Valentine's Day 1982. Entering his lab on a Sunday afternoon the young physicist made a heart-stopping discovery. His custom-built detector had just sensed something nobody had seen before — a particle called a magnetic monopole. For three years, Cabrera had been fine-tuning his experiment at Stanford University, California, and it seemed as if his diligence had paid off. "It certainly looked exciting at the time," he recalls. But today, more than 20 years later, he has yet to see a second monopole.

Most scientists would go crazy waiting years for a result that might never materialize. But a small number of physicists are doing just that. They devote years and even decades to the search for hypothetical particles and forces.

What keeps them going is a devotion to fundamental physics, an obsession with the strange machines they build to achieve their goals, and an opportunity for independence from the giant accelerator groups that now

dominate much of particle physics. "It's like drugs: you get addicted and you can't get out of it," says Giorgio Gratta, another physicist at Stanford University who will soon begin searching for a kind of nuclear decay that happens roughly once every trillion trillion years. Despite funding setbacks and the occasional false alarm, many of these researchers establish successful careers even when their detectors fail to record a single particle.

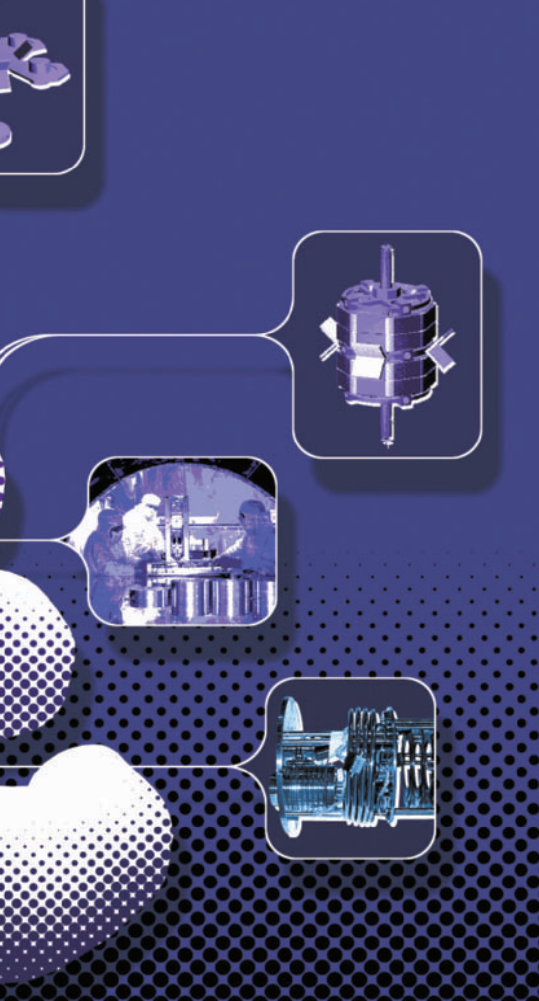
Part of the reason they command respect is because their quests address nagging questions left over from the past century's revolutions in physics. Einstein's general theory of relativity explained gravity's inner workings, and quantum mechanics led to a standard model that could predict the interactions of tiny particles, such as quarks — but try combining them and you get mathematical gibberish. A similar mismatch arises between cosmological theories and laboratory observations that fail to detect the

matter thought to hold together much of the Universe. These inconsistencies leave physicists feeling they are missing a big part of the picture.

To fill the gaps, theorists have generated a stream of hypothetical particles and forces. But these notions face a problem: if they were easy to find, somebody would have seen them already. So experimental physicists must go to extraordinary lengths to find them. They must build supersensitive detectors and shield them from the noisy world, they must ensure they do not fool themselves into seeing something that's not there — and above all, they must be patient.

Eric Adelberger, a physicist at the University of Washington in Seattle, could teach patience to a saint. He has spent the past decade watching moving pendulums in his lab. Adelberger's group uses the pendulums to test the behaviour of gravity at very small scales — where Einstein's laws conflict with

"Despite funding setbacks and the occasional false alarm, many researchers establish successful careers even when their detectors fail to record a single particle."



events. One wet November, a graduate student found that the gravitational field around the pendulums seemed to be growing inexplicably day by day. "It was the rain soaking into the Earth and changing the gravity field in the lab," Adelberger says.

"The difference in gravitational fields we can measure now is equivalent to the difference between a certain point in front of my nose and a point 1.6 nanometres higher," he says proudly. But in over ten years, Adelberger's probes have revealed little more than Seattle rain.

That doesn't bother him in the least, and it hasn't hurt his career. Adelberger became a Fellow of the American Physical Society in 1984 and was elected into the National Academy of Sciences in 1994. He credits his success to the effect his pendulums have had on theorists' ideas. "The negative results we are producing have a real impact on physics," he claims.

Non-event

The even greater impact that a positive result would have means that workers on these experiments must guard against the risk of self-delusion. It's a problem that has preoccupied Rainer Weiss, a physicist at the Massachusetts Institute of Technology in Cambridge. Weiss has devoted his career to detecting gravity waves — tiny ripples in the Universe's gravitational field that would have huge implications for cosmology if they could be seen.

That was what Joseph Weber, a physicist at the University of Maryland, College Park, claimed to have done in the late 1960s. Weber said that gravitational waves striking a giant bar of aluminium in his laboratory caused it to ring like a bell. But critics pointed out that the power needed to generate these waves would wipe out the entire Milky Way. Weber, it seemed, had been fooled by his own statistical analysis, and the controversy set back the field of gravity waves for decades.

Last year, the field entered the mainstream, when the \$300 million Laser Interferometer Gravitational-Wave Observatory (LIGO), a pair of detectors in Louisiana and Washington state, fired up giant microwave lasers that Weiss hopes will be able to spot gravity waves coming from the nearby Milky Way. With millions of dollars and the careers of some 400 researchers on the line, the group knows it cannot afford to cry wolf. "We have a set of physical conditions that have to be met for us to take an event seriously," Weiss says. But noise — such as thermal and mechanical vibrations — could blur the signal and trigger events that could be real or just an illusion. "You will then find us playing the same game that Blas Cabrera did," Weiss says.

Cabrera's Valentine's Day detection left him with a tough decision. Magnets have a north and a south, but for decades some theorists have predicted that a few lone norths or souths might wander the cosmos. These could account for a big part of the Universe's missing matter. They would also make physicists rethink the standard model, which does not predict monopoles.

But Cabrera knew his discovery was problematic. Although it matched theoretical predictions, and there were no obvious experimental errors that could explain the result, it was only a single event — hardly enough to overthrow established thinking.

Dark materials

In the end, Cabrera wrote a paper that was published in the prestigious journal *Physical Review Letters* (B. Cabrera *Phys. Rev. Lett.* **48**, 1378–1381; 1982). He didn't claim to have found monopoles, he just described the detector and the event. "Many of us said: 'My God, why the hell did he do that?'" says Weiss. But Cabrera's colleagues came to agree with his approach. "He published what he had, showing exactly what he did," Weiss says. He has invited Cabrera to talk to the LIGO group about what to do if they find an uncertain event.

Cabrera's device never spotted another monopole. His group has built second- and third-generation detectors, the last of which was thousands of times more sensitive than the original. But they came up empty-handed. "We never again saw an event like this one in any of the individual instruments," he says. "It seemed less and less likely that the one event we saw was a monopole."

But the negative results were not fruitless. They challenged theorists, including Alan Guth and Henry Tye, then at the Stanford Linear Accelerator Center, to explain monopoles' rarity. This influenced Guth's work on inflation theory — the idea that the Universe expanded exponentially just after the Big Bang. Astronomical observations have since confirmed this, and the theory has become one of the pillars of modern cosmology.

In the early 1990s, Cabrera switched to seeking other particles that might explain the Universe's missing mass. The resulting Cryogenic Dark Matter Search has run for 14 years: the latest detector, shielded from radiation deep within a Minnesota mine, was switched on last year. Cabrera is unconcerned that so far they have found no new particles. "If they're there, we will see them," he says.

Geoff Brumfiel is *Nature's* Washington physical sciences correspondent.



quantum mechanics. "There are lots of ideas predicting new forces that would show up in these experiments," he says.

Rather than swinging back and forth, these pendulums twist as Earth spins. In one system, otherwise identical weights made of lead and copper are tested to see if gravity pulls on them differently. If it did, it would contradict Galileo's legendary finding that two balls of different material fall to the ground at the same rate. It could also point to new physics governing gravity on the smallest scales. For example, string theory, one possible link between gravity and the quantum world, predicts particles that would change the laws of gravitation over short distances.

Small is detectable

Before Adelberger got into gravity he was in nuclear physics, where experiments involve hundreds of researchers. His work on pendulums has given him individuality, he says. "With these kinds of experiments, you're not a little part of a big thing, it's really yours," he says.

The pendulums took years of patient refinement: "It's kind of like a child," he says. The test weights are electrically inert, and tonnes of lead are used to shield them from the gravitational fields of nearby objects, such as hills and buildings.

Even with careful planning, Adelberger's group has had to cope with unexpected

Crunch time for Korea's cloners

A team in Seoul has stolen a march with its work towards human therapeutic cloning. The researchers have been fêted, but an ethical controversy may threaten their work. David Cyranoski investigates.

Woo Suk Hwang's laboratory, set in the hilly campus of Seoul National University, is a cloning factory. In one room, a gaggle of blue-coated researchers sits around a table plucking egg cells from cows' ovaries. Next door, other members of the team dart around the lab's 12 micromanipulators. At one of these workstations, researchers punch holes in the cow eggs and squeeze out their nuclei. At the next, colleagues take the stripped-down eggs and fuse them with cells from adult cows.

This production line underpins a project to create genetically modified cows that should be resistant to mad cow disease¹. It is just one of several cloning projects that require Hwang's army of researchers, students and technicians to handle up to 600 pig, cow and dog eggs every day. Hwang, a veterinary scientist, moved into cloning research in the late 1990s because of its promise in producing livestock with precisely defined genetic characteristics.

But it was a cloning experiment involving human eggs and cells that catapulted Hwang to fame this February. His team, together with researchers led by his Seoul National University colleague Shin Yong Moon, announced that they had obtained embryonic stem cells from a cloned human embryo². This provided a proof of principle for therapeutic cloning — the idea of repairing damaged or diseased tissues using cells derived from the patient, which shouldn't be rejected by the immune system.

Therapeutic cloning is a controversial technique, opposed by anti-abortion



Copy shop: members of the Seoul team extract cow eggs from ovaries for use in cloning experiments.

activists because it involves the destruction of several-day-old human embryos. But additional ethical questions are swirling around Hwang's work, relating to the recruitment of the women who donated eggs for the project, and whether guidelines designed to protect their rights were correctly applied (see News, page 3).

A-list celebrity

When *Nature* visited Hwang's lab last month, however, he was riding high. Hwang has been a celebrity in South Korea since 1999, when he unveiled the country's first cloned cow. Since his February paper, he has been fêted both at home and abroad. On 20 April, Hwang received the South Korean government's 'best scientist' award, worth US\$260,000. Days later, he and Moon were listed in *Time* magazine's "A-list of the world's most influential people".

Interviewed in his office, Hwang sat on the arm of a sofa, seeming flustered but apparently enjoying the attention. He had just received a delegation from the European Union; later, guests from Germany were due. "Everyone wants to know why we did the experiment or if they can collaborate," he told *Nature*. In demand to give lectures worldwide, Hwang's itinerary for this year includes visits to Britain, Germany, Italy and Spain.

His research schedule is similarly busy. Hwang wants to clone animals for various agricultural applications. He is working on a

project to clone the endangered Siberian tiger (*Panthera tigris altaica*), long extinct in the wild in South Korea, by fusing cells from captive adults with cow, pig, and dog eggs. And he plans to clone pigs to provide organs for transplants to humans. These will be genetically engineered to minimize the chances of the rejection by the human immune system.

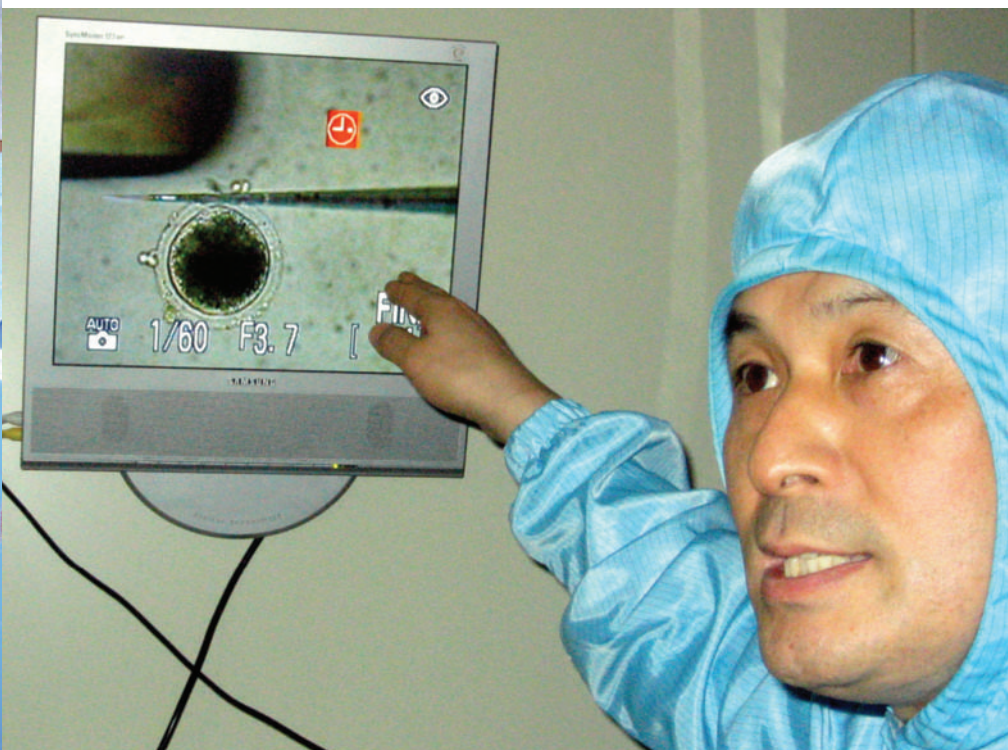
The human therapeutic cloning project also needs much work. "It's the very beginning," says Moon. "Please don't talk about clinical applications." Before therapeutic cloning becomes a reality, the production of

cloned embryonic stem cells must be made more efficient, as must the techniques of growing them into specific cell types to treat conditions such as Parkinson's disease, diabetes, or spinal cord injury.

Hwang and Moon made their cloned embryos by fusing eggs that had been stripped of their nuclei with cumulus cells. These cells, which surround and nurture developing eggs, came from the same individuals who donated the eggs. Attempts to clone embryos by fusing cumulus cells from one woman with eggs from another have so far failed. For therapeutic cloning to reach the clinic, it will be necessary to obtain viable embryonic stem cells using various types of cells from both men and women.

The researchers also want to see whether it will be possible to avoid the need for human egg donors by using cow eggs instead. Other groups are working along similar

"The cultures made from cloned cells are so precious that they are held in three separate, secure locations that are closed to visitors."



G. G. YOUNG/JOONGANGILBO/EPN/AP/T. S. WARREN

Cell mates: their creation of cloned human embryos (below) has made Hwang Woo Suk (left) and Shin Young Moon famous in Korea and beyond.

lines. Last year, for instance, Huizhen Sheng of Shanghai Second Medical University in China reported that her team had obtained embryonic stem cells from cloned embryos created by fusing human cells with rabbit eggs³. So far, Hwang says that his most successful experiment has brought 9% of the human-cow hybrid embryos to the stage at which embryonic stem cells can be harvested. But they have not yet been able to derive a stem-cell line.

Magic touch

Hwang attributes the Korean team's progress to its members' "magic hands". Experts who have visited the Seoul campus also point to the sheer scale of its cloning facilities. "I've never seen anything like it," says Jose Cibelli of Michigan State University in East Lansing, who worked briefly with Hwang's group on its landmark human cloning paper. Cibelli was also struck by the Korean researchers' dedication and industry. Hwang says that members of his group commonly work from six in the morning until ten at night. "It is tiring, but they do not refuse," he says.

Hwang has some 40 researchers in his team at Seoul National University's College of Veterinary Medicine; their equipment, provided through generous government support, is state-of-the-art. The human cloning work, meanwhile, was conducted in a separate facility, which Hwang says was funded through private donations. Because some Koreans oppose the creation of human embryos to obtain stem cells on moral grounds, Hwang says he kept his therapeutic



cloning work separate from his state-funded projects on cloning animals.

At the nearby College of Medicine, Moon has 30 more researchers working on the establishment and characterization of stem-cell lines. Elsewhere on the Seoul National University campus, affiliated groups are working on every step of the process of creating cell-based therapies that might one day be used to repair faulty organs or tissues. In total, the university hosts some 180 researchers working on cloning and related aspects of regenerative medicine.

But perhaps the biggest factor underlying Hwang and Moon's success in their human cloning experiments was the recruitment of 16 women who were prepared to have hormone injections to make them super-ovulate. These donors provided the 242 egg cells that were required to obtain the single line of cloned human embryonic stem cells described in their paper, published in *Science*.

The resulting cell cultures are so precious that they are held in three separate, secure locations that are closed to visitors. "Even some of our research staff are not allowed to

go in there," says Hwang.

Keeping the embryonic stem cells themselves under wraps is understandable. But when *Nature* visited Seoul, citizens' rights activists and bioethicists were starting to complain about a lack of transparency surrounding the recruitment of the egg donors, and raising questions about how rigorously Hwang and his colleagues followed the ethical guidelines laid down for their research.

Egg collecting

These concerns will be heightened by an interview that *Nature* conducted with one of Hwang's PhD students, who was also a co-author on the paper, in which she said that she and one other member of the lab had donated eggs. The student later changed her story, denying she had donated eggs and blaming her poor English for a misunderstanding.

Bioethicists argue that such a donation would be a breach of good practice, which should keep donors and researchers at arm's length, so that principal investigators cannot influence donors directly. It would also apparently breach the guidelines adopted for the experiment, which should have excluded donors who might benefit from the research.

No other cloning research group has recruited so many egg donors. Women usually have the procedure either to conceive through *in vitro* fertilization, or in rare cases to donate eggs to other women. In the United States, women who donate eggs can be paid several thousand dollars for expenses and as compensation for the inconvenience of the

UPI/W. S. HWANG/SNU



A herd of clones? One aim of Hwang Woo Suk's cow-cloning programme is to create genetically modified cattle that are resistant to mad cow disease.

invasive medical procedures. When Cibelli experimented with human cloning in 2001, while employed by the company Advanced Cell Technology in Worcester, Massachusetts, he used less than 20 eggs⁴, obtained from donors who were paid in this way.

By contrast, Hwang says that the Korean egg donors were not paid, and were motivated by a desire to help sick people, and through national pride. Cultural differences may also partly explain the Korean team's success in recruiting willing volunteers: in Asian societies a greater stress is placed on serving the common good.

Hwang has worked hard to engage the Korean public with his research, giving several public lectures each week. He wants to promote South Korea's economic development, borne in part from his memories of a childhood marked by poverty. "I want to make Korea a more well-equipped country, especially in science and technology," says Hwang.

Public spirit

The motivations of the PhD student who initially told *Nature* that she was among the donors fit with this picture of altruism and intense patriotism. In her original interview, she mentioned a desire to help sick children, and her love for Korea.

Challenged about his PhD student's initial statement, Hwang said he checked the informed-consent forms signed by all 16 women and could not find her name. Some

students did offer to donate eggs, Hwang added, but he "strongly refused".

Even before *Nature's* enquiries, questions were being raised about the recruitment of egg donors for Hwang and Moon's research. Experts expressed surprise that the team had been able to recruit so many women prepared to donate their eggs for free. In late March, the Korean Bioethics Association convened a committee to consider the issue. "It is all speculation now, but we want to investigate whether there were any vulnerable donors involved, maybe some young female researchers or women with a conflict of interest," says the association's secretary, Young-Mo Koo, a medical ethicist at the University of Ulsan College of Medicine.

The Korean Bioethics Association is pushing the National Human Rights Commission to investigate several aspects of Hwang and Moon's research, including the recruitment of egg donors, and whether the local Institutional Review Board that gave ethical approval for the project did an adequate job. The government-funded commission is expected to decide whether to investigate by 21 May. "It is a very delicate matter. Dr Hwang's experiment is a very big issue in Korean society," says Young Jun Choi, an expert adviser to the commission.

This sensitivity is underlined by the reactions of some Korean scientists who admit to concerns about the ethical questions surrounding the research. "No one wants to debate the ethics because the government is

so excited about it," says one senior biologist at Seoul National University. "Most scientists are also worried about a lack of students in science, so they don't want to break the excitement either. We need a hero."

Under pressure

For now, Hwang's work on human cloning is on hold. A national law was approved last December that would require a government committee to provide a license before any research using human eggs can be done. While the law does not come into effect until January 2005, Hwang says he wants to wait until the licensing system is in place. "If we can reach a consensus, we can continue our research with the support of the whole society," he says.

But Hwang's critics argue that the research reported in his *Science* paper has already moved beyond the national consensus. "He did not have the social acceptance of the people," claims Jae-kak Han, who heads the scientific division of People's Solidarity for Participatory Democracy, one of South Korea's leading citizens' rights groups.

Given the questions that are now being raised, the pressure is on Hwang to demonstrate that his work followed appropriate ethical guidelines. Any suggestion that it did not could undermine both Hwang's own plans to press ahead with therapeutic cloning research, and provide ammunition for critics of the technique worldwide. ■

David Cyranoski is *Nature's* Asian-Pacific correspondent.

1. Cyranoski, D. *Nature* 426, 743 (2003).
2. Hwang, W. S. et al. *Science* 303, 1669–1674 (2004).
3. Chen, Y. et al. *Cell Res.* 13, 251–263 (2003).
4. Cibelli, J. B. et al. *J. Regen. Med.* 2, 25–31 (2001).

Managing fisheries in a changing climate

No need to wait for more information: industrialized fishing is already wiping out stocks.

Sir—Your News story “Climate findings let fishermen off the hook” (*Nature* 428, 4; 2004) offers the view that climate warming and other environmental changes may be just as important as overfishing in driving worldwide declines of fish stocks. It also suggests that climate may be responsible when depleted stocks fail to recover, and that fishery management and climate research should be more closely tied to each other.

While it has been known for many years that climate variability can affect fish recruitment (the replenishment of stocks with juveniles), especially at the margins of species' ranges, there is at present no evidence that worldwide declines are linked in any major way to climate change. Large declines in marine fish communities have always coincided precisely with the onset of industrialized fishing (see R. A. Myers & B. Worm,

Nature 423, 280–283; 2003), which occurred at different times in different regions. Likewise, historical declines of marine megafauna, such as large groundfish, turtles or mammals, have always been related to the introduction of new fishing techniques and increased exploitation by European colonists (J. B. C. Jackson *et al.* *Science* 293, 629–637; 2001), not to climate change.

Spectacular collapses such as those of Eastern Canadian cod stocks, which at first were attributed to climate change, turned out on closer analysis to be solely related to overfishing (R. A. Myers *et al.* *Ecol. Appl.* 7, 91–106; 1997).

We emphasize that heavily overfished stocks may be more sensitive to climate variability, because loss of diversity among locally adapted populations has impaired resilience. Indeed, these findings do not let fishermen (or fishery managers) “off

the hook”, but underline the need for more conservative management.

The abundance of spawning and juvenile fish is estimated for most commercial fisheries each year using scientific surveys and fisheries data. This information already integrates actual climate influences and can thus be used to determine optimal fishing quotas.

The problem is that scientific recommendations are almost always exceeded, and overfishing is allowed to continue unabated, despite better knowledge. This problem is a political one. There is already sufficient information available to manage global fisheries sustainably, with or without any additional understanding of climate change.

Boris Worm, Ransom A. Myers

Biology Department, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

Colourful history of Japan's rat resources

Sir—As passionate rat researchers we were delighted to see the rat on the cover of the 1 April 2004 issue of *Nature*. It is little known that Japan has a long history of rat research, including studies on coat-colour inheritance (see picture) that date back to the eighteenth century.

Chingansodategusa, a Japanese guidebook on the breeding of fancy rats and mice, was published in 1787, during the Edo period, when Japan's civilization was isolated from the rest of the world. Japan is still an island but, we hope, is not so isolated today.

Your News Feature “The Renaissance rat” (*Nature* 428, 464–466; 2004) hailed a remarkable rodent, but it did not mention a modern Japanese contribution to its research: the National Bio Resource Project for the Rat (www.anim.med.kyoto-u.ac.jp/nbr). This rat-strain repository, one of the largest in the world, currently has more than 200 rat strains, which are freely distributed — at no cost except shipping — to any interested rat researcher worldwide.

In addition, we also provide a database for more than 100 different phenotype parameters of the rat strains. This allows for strain ranking, a method by which the parameters are sorted and unexpected abnormal values in reference strains can be revealed. This project already covers almost 50 rat strains and will finally comprise 200. This sort of phenotype ranking has the



Rat race: illustrations from an eighteenth-century Japanese treatise on breeding rodents for colour.

potential to save rat researchers all over the world valuable time, money and animals.

Tadao Serikawa

Institute of Laboratory Animals, Graduate School of Medicine, Kyoto University, Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

Multiskilled mouse rivals Renaissance rat

Sir—The laboratory rat has obviously made immeasurable contributions to biomedical science and will continue to do so as long as humans conduct biological research. There are many considerations in the choice of a model organism, and rats certainly have their advantages, such as their relatively large size. But to suggest, as your News Feature “The Renaissance rat” (*Nature* 428, 464–466; 2004) does, that the rat is superior to the mouse for behavioural research is narrow-minded at best.

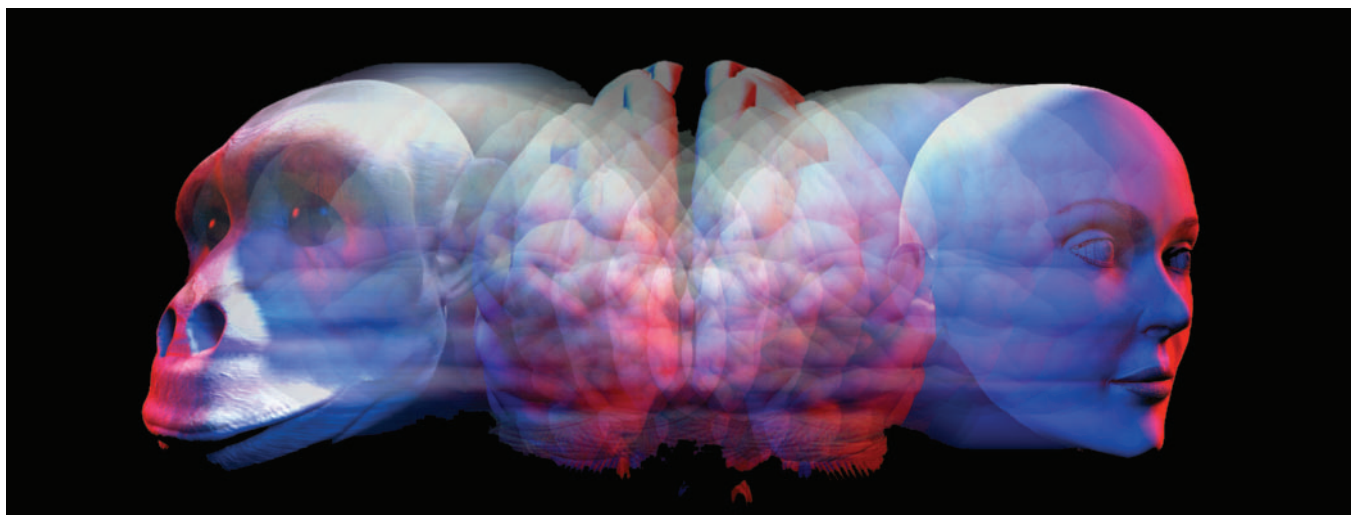
One only needs to work with a fairly

small number of the more than 2,800 inbred, recombinant inbred, congenic, consomic or transgenic strains currently available at the Jackson Laboratory in Maine (www.jax.org) and it soon becomes apparent that not all mice “just don't have a plan”, nor are all mice particularly aggressive. Mouse strains range from the very docile, such as A/J, to the very aggressive, such as NZB/B1NJ, and from the not-so-smart, such as LP/J, to strains that perform quite well on various tests of complex learning: C57BL6/J, for example.

Indeed, a similar range of responses can be found for any other behaviour of interest, and it is this amazingly diverse behavioural repertoire that has made the mouse extremely useful for studies of the genetic basis of variation in behaviour (see Crawley *et al.* *Psychopharmacology* 132, 107–124; 1997).

Richard A. Radcliffe

University of Colorado Health Sciences Center, Department of Pharmaceutical Sciences, 4200 East Ninth Avenue, Denver, Colorado 80262, USA



The human factor

What is it that makes us different from other animals?

So You Think You're Human? A Brief History of Humankind

by Felipe Fernández-Armesto
Oxford University Press: 2004. 208 pp.
£14.99, \$23

The Human Story: A New History of Mankind's Evolution

by Robin Dunbar
Faber & Faber: 2004. 224 pp. £12.99

David L. Hull

Writing for a popular audience is not quite as easy as authors generally think it is. Most popularizations are much too technical and not sufficiently engrossing. Both Felipe Fernández-Armesto and Robin Dunbar, on the other hand, have gauged their audience just right. They have simplified without falsifying and have picked a topic that people find endlessly fascinating — the nature of human nature.

These two books also complement each other nicely. Dunbar begins with the fossil record and then leaps ahead to the present. Most of his discussion concerns current ideas about human beings as we have studied them scientifically. Fernández-Armesto devotes most of his book to filling in the gap between our fossil ancestors and the present, expanding his investigations to include both the history and the prehistory of humans at a wide variety of places and times. Those of us living right now in the West assume there is such a thing as human nature that makes us different from other animals, and that this belief stretches back indefinitely in time. We might disagree about what this nature is, but we take for granted that it exists. As it turns out, a belief in human nature is of recent vintage and is not all that widespread.

In *So You Think You're Human?* Fernández-Armesto shows exactly how mistaken this

chauvinistic belief actually is. People in the West think that species can be distinguished from each other quite easily. But in the past, people in various cultures have thought that the boundaries between species are both permeable and fuzzy. In some cultures a wide variety of entities have been treated as if they should be classified with us. Conversely, some of those we regard as unproblematically human were in other societies not considered human at all. For example, when the neighbours of pygmies killed and ate them, they did not think they were doing anything different from killing and eating wild boar or monkeys. If they killed and ate each other, that would be cannibalism; doing the same to pygmies was not.

Although both Fernández-Armesto and Dunbar are interested in the same issue — human nature — they approach it differently. Fernández-Armesto overwhelms the reader with example after example, whereas Dunbar presents more careful, coherent arguments. Both authors discuss many of the same things: bodily differences such as the plantigrade (flat to the floor) foot, opposable thumbs, bipedal gait and big brains; mental differences in intelligence, speech, imagination, tool manufacture and use, fire and cooking; and finally social differences in culture, religion, music, art, reason, use of medicine, social learning and the formation of social groups.

Fernández-Armesto concludes that the differences between us and other primates are not differences of kind but of degree, albeit an enormous degree. Although Dunbar is sympathetic, he is forced to the opposite conclusion. The ultimate criterion for Dunbar is intentionality and an accompanying theory of mind. Until the age of four or so, human children have only a single level

of intentionality. They are aware of their own intentions, but they cannot distinguish between their own intentions and those of others. But after this crucial period, they can. As they mature, the levels of intentionality increase until they can understand that when Shakespeare wrote *Twelfth Night*, "he intended [1] that his audience should realise [2] that the much derided Malvolia believed [3] that his mistress Olivia wanted [4] to marry him instead of his being her servant."

Chimpanzees and autistic children are not able to develop a theory of mind. Chimpanzees are "on the brink," says Dunbar, but that is all. This dispute does not seem to afford any sort of resolution. One side sees only differences in degree. The other side sees a difference in kind, and if not in kind, at least, massive differences in degree.

Fernández-Armesto ends his book by concluding that we need to rethink how we conceive of the living world, in particular the human species. I could not agree more. In the recent past, at least, everyone who has addressed this question has assumed that species are kinds defined in terms of characteristics — the task was to find the right ones. But what if we do not treat species in this way at all but as individuals? Before the acceptance of evolutionary theory, species seemed to be paradigmatic kinds, but with the introduction of evolutionary theory, new dimensions were introduced: place and time. For selection to work, descent is required, and that is a spatiotemporal relation.

If we accept the role of variation and selection in the evolution of species, then descent must take priority over similarity. The problem is that we glide too readily over fundamentally different notions. For example, species have boundaries, but two quite different sorts of boundary are being

invoked, only one of which is spatiotemporal. Biogeographers trace the spatiotemporal boundaries of species, discovering for example that catalpa trees are indigenous to the Wabash Valley. The other sorts of boundaries are conceptual and concern definitions. All triangles, and only triangles, have three sides.

Although not all species reproduce sexually, those that do so produce a genealogical nexus, and species are chunks of it. At best, the way that characteristics are distributed is secondary to descent. In referring to species as the groupings that evolve by variation and selection, I do not mean to imply that this is the only way of defining species. Lots of

people do it in lots of different ways, but Fernández-Armesto and Dunbar both introduce Darwin and evolution. So they need to be made aware of one quite prevalent way of construing species, including humans, especially as it provides rather a different perspective on the human species. Fernández-Armesto devotes his final chapter to issues surrounding morality. Once again, as he sees it, the issue is characteristics. Which characteristics are relevant to morality, and which organisms exhibit them? What should we do with non-human organisms that have the characteristics that we use to confer moral rights?

From the evolutionary perspective, how-

ever, the issue is genealogy, not distributions of characteristics. Even though some pigs may seem brighter than some people, pigs do not belong to the same chunk of the genealogical nexus as people. Evolutionary theory is inherently species-ist. If we discovered that dolphins have sufficiently well-developed language skills that we could strike up conversations with them, they would still be dolphins and we would still be humans. Social and moral problems would arise, but as biological species we would remain unique and distinct. ■

David L. Hull is emeritus professor at the Department of Philosophy, Northwestern University, Evanston, Illinois 60208-1315, USA.

Walking on their ribs

Geoffroy Saint-Hilaire: A Visionary Naturalist

by Hervé Le Guyader,

transl. Marjorie Grene

University of Chicago Press: 2004. 302 pp.

\$45, £31.50

Nick Hopwood

In evolutionary terms, the most remarkable discoveries of developmental biology in the 1980s and 1990s were that the molecular mechanisms of anteroposterior axis formation are shared across most of the animal kingdom, and that vertebrates form a dorsoventral axis in a similar way to insects, only upside-down. This axial inversion had a special appeal because it seemed to confirm an old and much-derided view. Its first proponent was the French zoologist Étienne Geoffroy Saint-Hilaire, subject of this book by Hervé Le Guyader.

In 1830, Geoffroy faced his one-time friend and long-term colleague Georges Cuvier at the Muséum d'Histoire Naturelle in Paris, in one of the most famous controversies in the history of science. Cuvier, the most powerful comparative anatomist of the age, had divided the animal kingdom into four completely separate branches: vertebrates, articulates (largely arthropods and annelids), molluscs and radiates (echinoderms, cnidarians and various other groups). Even within these divisions, he allowed structural similarity to result solely from the same functional demands.

Geoffroy, by contrast, taught that function did not really matter, nor even form; what counted were the connections between the parts. He founded a 'philosophical' anatomy on 'analogy' (homology, to us), and pushed the idea that all animals are built to a single plan. Having established a common scheme for vertebrates, in 1820 he extended it to the articulates. Insects, he pronounced,



are vertebrates that live within their vertebral columns and walk on their ribs.

Cuvier restrained himself until a decade later, when Geoffroy exploited two young naturalists' suggestion that cuttlefish, representing molluscs, were like vertebrates doubled back on themselves. The simmering dispute now boiled over into a confrontation on the floor of the Academy of Sciences in Paris that captivated the learned world and newspaper-reading public alike.

As Toby Appel showed in *The Cuvier-Geoffroy Debate* (Oxford University Press, 1987), much more than cephalopod anatomy was at stake. Cuvier's functionalism opposed Geoffroy's morphology. The austere, logical Protestant fought the intuitive, impetuous, romantic Deist. Cuvier's fact-driven, establishment science was pitted against Geoffroy's broad speculation and alliances with progressive forces beyond the academy. Cuvier is usually said to have won, and on the narrow issue he did, but Geoffroy's philosophical anatomy was more influential than used to be thought. For leading naturalists in the next generations, both positions seemed too extreme. In finding a resolution, homology was made into evidence of darwinian evolution — and

some darwinists argued that the vertebrates originated from annelids by inversion of the dorsoventral axis.

Le Guyader's book, first published in French six years ago, offers a judicious selection of Geoffroy's works, each with a short introduction. These texts have never been translated before because nineteenth-century British naturalists read them in French, and because by the middle of the century Geoffroy had been marginalized. We have the preliminary discourses to both volumes of his *Anatomical Philosophy*; the first of three treatises on the organization of insects, which deals with the anteroposterior axes of insects and vertebrates; the recently much-cited review, "General considerations on the vertebra", with its reflections on dorsoventral organization and the figure of the inside-out and upside-down lobster; and the whole of the *Principles of Zoological Philosophy*, Geoffroy's record of the 1830 debate.

Anglophone biologists and historians of science will be glad to have these scarce and important works so readily available. Some additional editorial work might, though, have produced more scholarly and user-friendly texts. Martin Rudwick's translations in *Georges Cuvier, Fossil Bones and Geological*



Catastrophes, published by the University of Chicago Press in 1997, show the advantages of identifying authors mentioned, giving full references for cited works, explaining allusions and discussing tricky points. But if the English prose here rarely sparkles, Geoffroy is largely to blame. On style, Cuvier won hands down.

So why should we read these works today? The Cuvier–Geoffroy debate has been given many different meanings, by the two sides at the time and by commentators since. For Le Guyader, molecular developmental biology allows us to recognize Geoffroy as a “visionary of genius” who—Le Guyader realizes it would be grossly anachronistic to call him right—was “right to be wrong”. A paper from 1822 certainly generates additional interest now that it is cited again. But the deeper reasons why Geoffroy still matters are the approaches that he and Cuvier framed and fought over, rather than any specific answers he gave. Their views decisively shaped our science. ■

Nick Hopwood is in the Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge CB2 3RH, UK. He is the author of *Embryos in Wax: Models from the Ziegler Studio*.

Sexual diversity and the gender agenda

Evolution's Rainbow: Diversity, Gender and Sexuality in Nature and People

Joan Roughgarden

University of California Press: 2004. 472 pp. \$27.50, £18.95

Sarah Blaffer Hrdy

Rather than being one coherent book, the narrative of *Evolution's Rainbow* shuttles between three interwoven agendas. The first is a passionate cry from the heart for greater understanding of sexual diversity in nature and greater tolerance for the many gay men, lesbians, bisexuals, transgenders and others

who do not fit comfortably into male or female binary categories.

The author, eminent Stanford biologist Joan Roughgarden, who is herself a transgendered woman previously known as John, cites poignant case studies that illustrate the anguish leading up to decisions to switch genders. As one such transgendered woman put it: “I’ve laid everything I’ve achieved in life—job, relationship, family, health, future—on the table, and it seems fate will decide what I am allowed to keep, if anything. It’s kind of like starting life all over again.” Roughgarden even includes rare practical advice on how to inform your boss that you intend to switch genders without losing your job. Few readers of *Nature* will disagree with the main tenets of what is essentially a human-rights agenda, even if they don’t agree with all of Roughgarden’s interpretations or policy recommendations.

The second, and for me most interesting, book-within-a-book provides a cornucopia of information about sex and gender diversity across human societies and the natural world. There are vignettes about homosexuals in Ancient Greece, eunuchs in Rome, hijras in India, and Native American ‘two-spirits’. Roughgarden then takes us on a whirlwind tour through the zoological counterpart of human gender studies, introducing fish that change from females into males; intersex kangaroos that have a penis as well as a pouch equipped with mammary glands; kangaroo rats, in which up to 16% of a population have both sperm- and egg-related plumbing; intersex bears that give birth through the penis; hermaphroditic whales; and homosexual black swans that turn out to be more successful at rearing young than their heterosexual counterparts.

For those familiar with the literature on sex differences, some of these gender-bending examples will be old hat, others quite new. A few sections are dazzlingly original, such as Roughgarden’s imagined memoir of her own embryonic development. Wherever possible, she emphasizes the naturalness of gender gradations. She estimates that one in a thousand people are transgendered.

Iconoclastically, Roughgarden suggests that transsexuality is unlikely to be an anomaly: throughout the book, she is at pains to suggest ways in which gender ambiguity (such as feminine-seeming male fish) might be adaptive.

As a compendium of information on sex and gender diversity in the natural world, Roughgarden’s is the richest and most authoritative book available, fulfilling a desperate need of readers both within and outside the scientific community. For readers craving information about transgendered existences, or for those like me who are deeply moved by Gay Pride parades and the social transformations that they represent, this book is going to have a huge impact. This is why I wish the third agenda had been more carefully considered.

In what Roughgarden herself views as the main message of the book, she lays out her reasons for rejecting Darwin’s theory of sexual selection, as well as her grounds for an “indictment of academia for suppressing and denying diversity”. The “facts of nature”, she claims, falsify darwinian theory. She stresses instead that sex is essentially cooperative, is not a battle and often does not lead to conception. “We are only just realizing,” she writes, “that the concepts of gender and sexuality we grew up with are seriously flawed.” She argues that Darwin’s theory of sexual selection is so flawed that it should be replaced by a theory of her own. However, I am less impressed by her critiques and alternative theory than by her timing. This evolutionary biologist becomes a woman, and only then do the problems occur to her?

It is a bad idea, argues Roughgarden, to move from a binary gender system based on gamete size (large eggs and small sperm) to explanations about differences in male and female body sizes, behaviours or mating strategies. It is ill advised, she says, to stereotype males as sexually ardent and females as coy; to ignore social or ecological constraints on female choice; to overemphasize genes at the expense of social or developmental contexts; or to give precedence to quests for supposedly the ‘best’ genes at the

expense of non-conceptive strategies that animals might be pursuing. I could not agree more. But there is a Rip Van Winkle quality to all this.

Far from being “just realized”, these problems were discussed in the 1970s and 1980s by female sociobiologists, who called for a broadening of evolutionary theories to include selection pressures at more levels and on a broader range of individuals (such as females and juveniles). Roughgarden mentions some of these critiques in passing, but without exploring the impact that such findings are already having. Even though the push for revisions began on the margins of evolutionary theorizing, change (however inefficiently reached) came from within. This process is representative of how science works. Over the long haul, researchers (especially young ones) have less to gain from suppressing information about diversity than from making their reputations by exploring it.

Yes, it took a long time. But once the previously marginalized (Katherine Ralls, Mary Jane West-Eberhard, Patricia Adair Gowaty, Barbara Smuts and myself, to name only a few) had gained enough influence to be heard, evolutionary horizons began to broaden. Far from trying to suppress our

views, new generations of sociobiologists rushed past us to explore the expanding frontiers. Theories that were never so much false as incomplete and overextended are already beginning to be revised. Roughgarden's book will similarly galvanize research in the transsexual realm of the great continuum of diversity out there.

But I fear that her book may also mislead readers on two counts. First, competition between those of one sex for reproductive access to the other remains a robust explanatory framework, even though it is not the whole story. Second, I fear that Roughgarden's “indictment” will fuel widespread misunderstandings about the way in which scientists operate.

Unquestionably, many medical professionals were slow to understand what being transgendered means. Lives were sacrificed to dogmatic surgical and psychological interventions. Some hopelessly unreflective scientists continue to “sneak gender stereotypes into the primary scientific literature and corrupt its objectivity”, aided and abetted by sloppy popularizers. Raising the consciousness of darwinists is taking longer than ideally it might have. Nor have evolutionary biologists always been candid in

acknowledging either the sources of our biases or how such biases came to be corrected. Evolutionary biology does indeed have its share of die-hard dogmatists.

But the biases that Roughgarden highlights have less to do with false theories or cover-ups than with the occupational hazards of being human, of tending to see what we expect (or sometimes want) to see and incorporating prejudiced observations into our starting assumptions. Empathy (just whose interests we identify with) plays a much greater role in the biological and social sciences than is generally acknowledged. It is even possible that learning biases are built into human psychology, preventing us from noticing things counter to our own (including our gender's) interests. But these are not reasons to reject sexual-selection theory; they are merely reasons to encourage diversity among those applying it, the better to study nature's rainbows, sexual and otherwise. Hopefully, Roughgarden's book will be one more step towards that goal. ■

Sarah Blaffer Hrdy is professor emerita in the Department of Anthropology, University of California, Davis, California 95616, USA. Her most recent book is Mother Nature: A History of Mothers, Infants, and Natural Selection.

High points in geology

Devil in the Mountain: A Search for the Origin of the Andes

by Simon Lamb

Princeton University Press: 2004. 336 pp.
\$29.95, £19.95

David E. James

Devil in the Mountain is the fascinating story of geologist Simon Lamb's quest to understand how the high Andes formed. It includes a wealth of real-life, even harrowing, anecdotes of fieldwork, mixed with colourful descriptions of Bolivian culture. This engrossing and well-written book focuses on deciphering how the modern Andes, and high mountains in general, took shape on the surface of the Earth. Lamb parses the Andean story for us, piece by piece, as he came to understand the process. The reader travels with him by Toyota Land Cruiser through the Bolivian Andes as he sorts through the clues to their growth. We share with him his experiences in New Zealand and Scotland, his opportune conversations with fellow scientists, and a host of other disparate threads that became the building blocks of his understanding of how high mountains are formed.

The story unfolds in the central Andes, the highest and widest part of the Andean chain that runs the length of western South



America. Here the great volcanic peaks of the Western Cordillera and the tectonically folded and faulted ranges of the high Eastern Cordillera are split from one another by one of the world's greatest plateaux, the altiplano, situated an imposing 4,000 metres above sea level. Lamb staked out his territory in 1989 in the Eastern Cordillera of Bolivia, and he and his students embarked on a protracted study to understand what drove the Andes to such great heights.

Lamb begins by explaining how he came to study the Andes in the first place, and the origins of the book's title — a reference to the evil underground spirits that Bolivian miners believe decide the fate of each mine,

and which they appease with a statue and offerings at the mine entrance. Then we divert to northern Scotland where Lamb introduces the enigma not just of mountain building but of Earth's evolution itself, starting with the early scientific breakthroughs that revealed deep geologic time and thence along the tortuous path of scientific ideas to the dawn of plate tectonics. The narrative of Lamb's own journey of discovery is inextricably linked to plate tectonics, which lies at the heart of all understanding of the Andes.

Lamb's writing is engaging and clear, making for a thoroughly accessible book. He is at his best showing how geologists decipher the clues hidden in outcrops and the

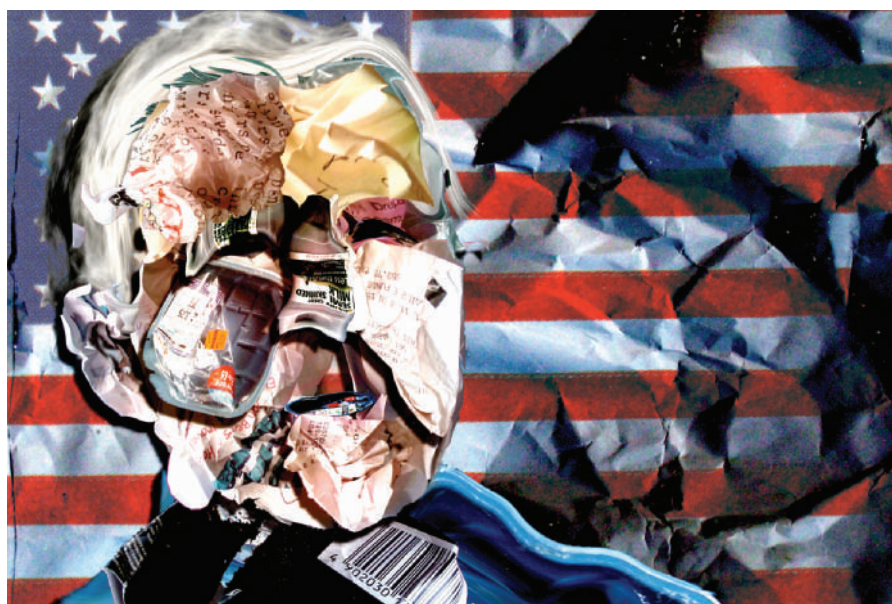
landscape. Even when he tackles esoteric areas of geophysics and dynamical modelling far from his own field of expertise, Lamb exhibits an enviable facility for simplifying complex and, in some cases, quite controversial ideas. He has keen insight into Bolivian geology and a sympathetic eye for local culture. His description of his first visit to La Paz was eerily reminiscent of my own experience some 20 years earlier — from the out-of-breath climb to the seismic Observatorio San Calixto and the indispensable meeting with its director, Father Cabré, to the struggles over drafting a formal agreement, or *convenio*, with the Bolivian Geological Survey.

Amid all the hazards and exhilaration of fieldwork in a distant land, Lamb succeeds in capturing Bolivia: the pervasive smell of diesel fumes mixed with an aroma of urine in the streets of La Paz; the “thirty-six signatures, several from the same official” required to get equipment through customs; the endless efforts to keep dilapidated vehicles running on atrocious roads at high altitude; the bewildering networks of rutted tracks in the altiplano that go nowhere; the omnipresent military checkpoints; and the almost unbelievably hard life of the average Bolivian miner or *campesino*.

As a popular account of Andean formation, *Devil in the Mountain* makes compelling reading. Lamb has done a masterful job in piecing together the Andean puzzle in a way that seems to make perfect sense. Nonetheless, from a strictly scientific point of view, his highly personalized take on the origins of the Andes contains numerous controversial ideas, and the scientific community remains split on many key questions. Lamb reveals throughout a strong Oxbridge-centrism (he is, after all, an Oxford lecturer with a PhD from Cambridge) that dominates many of the ‘big picture’ sections in the later chapters. Moreover, the same big picture that makes sense in one part of the Andes may not do so in another. To his credit, he is well aware of the unfortunate way in which research into the Andes has been compartmentalized over the years, and has worked harder than most to escape the provincial thinking that comes with it.

Only occasionally does he stumble outright, for example when he seemed unaware in 1995 that seismic and gravity studies had shown 25 years previously that a deep root beneath the Andes compensated for the high elevations. Overall, however, these few reservations seem to be relatively minor grains of salt to swallow. Despite many personal quibbles of my own, I thoroughly enjoyed this book, which in the end is a fair and plausible account of how high mountains are born and how they die — and even, perhaps, how they affect the life of the planet. ■

David E. James, is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC 20015, USA.



The way or the world

One with Nineveh: Politics, Consumption and the Human Future

by Paul R. Ehrlich & Anne H. Ehrlich
Island Press: 2004. 430 pp. \$27

Norman Myers

What, you might say, yet another book from Paul and Anne Ehrlich on the environment, population and consumption, and about how humans are grossly degrading Earth's life-support systems? Can the Ehrlichs have anything new and different to offer about the crises they have been trumpeting for decades? Well, they certainly do in this latest book, which is full of pioneering analyses and innovative insights.

The book's initial purpose is to present our predicament in terms of the environmental ruin, fostered by political hubris and citizens' myopia, that overtook the ancient Assyrian capital Nineveh. The disaster was poetically exemplified by Rudyard Kipling's “Lo, all our pomp of yesterday/Is one with Nineveh and Tyre.”

After this stage-setting, the Ehrlichs review our dire environmental prospect, focusing on shortages of food and energy, biodepletion, loss of ecosystem services, and grand-scale pollution, among a litany of related ills. Then they assert that there are two principal drivers of our predicament: too many people and too much consumption. All these topics are dealt with in detail and with evidence piled on evidence. Certain sections could have been compressed, as they have been rehearsed on many occasions already — even if not everyone has been listening.

Parts of the book are a vigorous critique

of US President George W. Bush and his policies — and, by extension, of the ultra-right community that seems to dominate American society. Bush often reiterates his father's thesis, as enunciated, for example, at the Rio Earth Summit in 1992 when Bush senior declared, with reference to his country's fixation on fossil fuels: “The American way of life is not negotiable.” Both presidents Bush have asserted that a shift away from fossil fuels would knock deep dents in the US economy. And both have failed to see that it is precisely the fixation on fossil fuels that will, courtesy of global warming, transform the American way of life from top to bottom. I suspect that people of the future will not regard today's spot price of oil or the latest sales figures for sports utility vehicles as the key statistics of our time, rather than the United States, with 4.6% of the world's population, produces 24% of the world's carbon dioxide emissions.

The Ehrlichs deal several times with the absurdity, if not the arrogance, of the presidential proclamations, pointing out that not even the United States, however powerful it may be, can insulate itself from links with the climate. The winds carry no passports and Americans are becoming, first and foremost, citizens of the global community, even though many of them prefer to proclaim their ‘exceptionalism’. They epitomize the dictum that, first, no country can support an indefinite increase either in its number of people or in its consumption of environmental resources, let alone both; and second, most mainstream policies of most governments assume, on the contrary, that they can. This is true of most nations, to a degree, but the United States is in a league of its own.

The environmental crisis has been well covered in the Ehrlichs' earlier books, so what is new and different about this one? Here the Ehrlichs offer us much more on the hard-nosed solutions needed to expand

society's understanding of our environmental travails and the institutional inertia that ensues from ignorance (or ignore-ance). Instead of lengthy lists of obvious safeguard measures such as switching off unnecessary lights (even while leaving computers and microwaves on standby), the Ehrlichs take us through the systems, political and otherwise, that we deploy to run our societies. They highlight the cultural disconnection between what most people think is going on (the stock market, say) and the environmental breakdowns that will eventually prove much more important. The technological underpinnings of our societies have far outstripped our cultural understanding. "A 747 jetliner embodies much more information...than all of the DNA packed into the pilots' cells...The transition to a state of ubiquitous cultural ignorance has occurred in an evolutionary twinkling of an eye, and humanity is having great difficulty dealing with it." This applies particularly to those technologies that promise much but inflict severe harm on the human cause.

The book is not only about the usual environmental problems. It tackles sources of problems, such as cultural roadblocks, group (mis-)behaviour, contagion of attitudes, "stickiness" of thinking, social discontinuities, special interests and other forms of social rigidity and perversity. The Ehrlichs also present plenty of solutions through their assessments of collective action, optimized governance, pre-emptive initiatives, ethical imperatives and huge changes in personal and community outlooks. This material makes up the bulk of the book, in contrast with the Ehrlichs' earlier writings. Here they deal with issues that deserve intensive analysis but receive comparatively little research.

Not surprisingly, the Ehrlichs suggest that the Millennium Development Goals proposed by the United Nations in 2000—to tackle malnutrition, child mortality, female illiteracy and water shortages, for example—be supplemented by a Millennium Assessment of Human Behaviour. Such a measure could be founded on the idea that we live at a time when what is idealistic is swiftly converging with what is realistic.

This listing of the principal components of the Ehrlichs' message may sound a trifle earnest, but the book's spirit is saved by the jaunty style that informs the text. The authors are masters at finishing paragraphs with one-liners, such as: "We have utterly changed our world; now we'll have to see if we can change our ways." Better still, the style suggests that the authors are not a whit dispirited by their message; rather, it implies that they sense not only profound problems but superb opportunities. The Ehrlichs have often been called the ultimate pessimists, but their book is, frankly, heartening.

In short, they have spread their conceptual net much wider than before. The book is

decidedly new and different. And as a final bonus, it contains more than 700 references and 1,000 notes. ■

Norman Myers is honorary visiting fellow at Green College, Oxford University, Upper Meadow, Old Road, Oxford OX3 8SZ, UK.

Family values

More than Kin and Less than Kind: the Evolution of Family Conflict

by Douglas W. Mock
*Harvard University Press: 2004. 352 pp.
\$27.95*

H. Charles J. Godfrey

The private life of the Verreaux's (or black) eagle, a spectacular raptor of the drier parts of Africa, does not exactly embody Victorian family values. The female nearly always lays two eggs, a few days apart, which hatch into two chicks, one a little bigger than the other. As soon as the smaller hatches it is turned upon by its sibling, which attacks it viciously, the assault almost invariably leading to death. In one nest that was continuously watched by the ornithologist Valerie Gargett, the bigger chick pecked its nest-mate 1,569 times in its brief 72-hour existence. Throughout this period, the parents did nothing to intervene. Ornithologists used to call this phenomenon 'cainism' but today use the less poetic, but less sexist, term 'siblicide'. Obligate siblicide occurs in some other raptors as well, and in species of boobies, penguins, egrets and pelicans. Less extreme forms of siblicide are much more widespread in birds: usually the parent produces a hierarchy of chicks and, when conditions are less than perfect, the runt dies, either because its siblings passively deprive of it food, or through more direct attacks.

In *More than Kin and Less than Kind*, Douglas Mock surveys sibling conflict in birds, and gives a more limited number of examples from other groups of animals and plants. He charts how biologists' views of siblicide have evolved over the years: it has been considered simply an aberration, and also a means of population control for the benefit of the species. The latter explanation became untenable when evolutionists in the 1960s

realized the importance of thinking about how selection operates at the individual level. David Lack's view prevailed: parents overproduce young so that they can capitalize on bumper food years when everyone survives, with siblicide being an efficient mechanism to adjust litter size in poor years.

This still does not explain obligate siblicide, which for a long time was controversial. But now there is a general consensus that obligate siblicide is an insurance mechanism to guard against egg infertility or the very early death of a chick: in the species in which it occurs, an infertile single egg may mean a completely wasted breeding season.

Although extreme forms of sibling conflict receive most attention, Mock also discusses more moderate forms of squabbling (informed, perhaps, by his own experience as the youngest of four brothers), as well as conflicts between parents and offspring, and between mothers and fathers.

Mock has a lively and engaging style, and the skill to explain complex ideas from kin selection and related fields intelligibly without being patronizing. He is best known for his work on siblicide in egrets and herons, and his accounts on the one hand of long hours of fieldwork with the birds steadfastly refusing to behave as his hypotheses predict, and on the other of sudden insights from watching family interactions in the nest, are both inspiring and ring true.

The one area where I think he goes astray is parent-offspring conflict, which he views as an interesting theory that, with a few notable exceptions, has failed to produce much insight. The theory is based on Robert Trivers' argument that natural selection operates differently on genes expressed in





offspring and parents. Mock argues that in a “really compelling case” of parent–offspring conflict, selection would produce the optimum outcome for offspring with a corresponding reduction in parental fitness. I think this is far too narrow an interpretation, and results from a failure to distinguish between the genetic battleground identified by Trivers, and the resolution of the conflict that we see in nature.

For example, it has been argued that the presence of parent–offspring conflict means that the only evolutionarily stable way in which young can communicate their requirements for food to their parents is by using costly signals. This would explain why the begging calls of baby birds may be energetically expensive to produce, or risk attracting predators. If true, the theory has told us something new about a situation where the optimum outcome for parents is achieved, though at a cost. Mock says little about this type of honest signalling by offspring, and in places is, I think, unfairly dismissive. This is a shame, as these ideas can explain many of the same phenomena as sibling conflict, and the two processes almost certainly operate together. Behavioural ecologists such as Mock have pointed to the difficulty of testing honest-signalling theory, but I think this should be viewed as a challenge rather than a reason to dismiss it.

With this one caveat, Mock has done a superb job in bringing a large area of contemporary behavioural ecology to both a biological and a general audience. It will stimulate a new generation of fieldworkers and inspire theoreticians to try to understand the complex interplay of natural selection acting simultaneously on different members of the family. For a general reader, it will explain how modern evolutionary theory is tested in the field and give a real flavour of the life a field biologist. It deserves to be read by everyone interested in the evolution of family life. ■

Charles Godfray is at the NERC Centre for Population Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK.

Warming to a historical theme

The Long Summer: How Climate Changed Civilization

by Brian Fagan

Granta/Basic Books: 2004. 284pp. £20/\$26

Jeremy A. Sabloff

Archaeology has numerous goals, which include constructing histories of peoples' cultures through space and time, offering an appreciation of the achievements of past civilizations, and providing historical contexts, both cultural and ecological, for modern events and processes. Writing about such issues for the public can be a challenge, but Brian Fagan is one of the world's most accomplished and prolific popularizers of this complex discipline. As he has shown in a variety of books, he has the knack of making arcane archaeological information accessible to broad audiences. Coupled with his broad knowledge of world prehistory, this skill allows him to show how the distant past is relevant to current concerns.

One of the most pressing concerns today is global climate change. In *The Long Summer*, Fagan provides an enlightening context for this key modern problem through a clear-headed discussion of the past 15,000 years of climatic warming, which has been enriched by great advances in climatology. He goes on to show how this climatic trend has influenced, but not determined, the development of complex civilizations.

Fagan states that “human relationships to the natural environment and short-term climatic change have always been in flux”, and that we have become increasingly vulnerable to the effects of climate change, thanks to population growth, urbanization and the global spread of industry. He argues that our attempts “to cushion ourselves against smaller, frequent climate stresses...have consistently made us more vulnerable to rarer but larger catastrophes”. This argument

leads him to conclude that: “the present problem of global warming is neither proof of late capitalism's intent to commit industrial-strength sins against Mother Earth nor a hallucination imposed on the world by anti-business activists. It is simply a reflection of the scale of our vulnerability, the scale on which we must now think and act.”

The environmental climate trajectory that Fagan describes is not a simple one; it has numerous short and long perturbations. The number of places, environmental events and cultures discussed can be daunting and occasionally confusing, and the complex story sometimes seems to get the better of even such a skilled storyteller as Fagan. But perseverance pays off, as Fagan manages to draw the diverse pieces into a coherent narrative and keep the reader on track. Moving forwards in time, Fagan examines the development of cultural complexity in selected time periods from 15,000 years ago to the present day, and in different regions of the globe from the Near East to the Andes.

Fagan offers many examples of how changes in climate have influenced cultures from hunters and gatherers to complex civilizations. One of the common factors is that environmental shifts in one region of the world can have profound effects in distant regions of the globe. For instance, the collapse of the Laurentide glacier in northern Canada around 6200 BC had profound effects in the eastern Mediterranean and Anatolia, even triggering a drought lasting four centuries, which in turn had a major cultural impact on Anatolia.

Fagan also offers gripping examples of the key role of climatic change in such diverse topics as the rise of settled villages in the Near East some 11,000 years ago, the early growth of cities in ancient Sumer and Egypt in the fourth millennium BC, and the fall of the Roman Empire by the end of the fifth century AD.

Experts in different areas may take issue with Fagan over the details of the changes in the regions that they know best. Clearly Fagan has had to paint with a broad brush,

but even though scholars might nitpick on the specifics, the overall picture he creates is argued convincingly.

Fagan concludes the book with a brief review of the evidence that recent years have seen accelerated warming. "With the Industrial Revolution, we took a great stride into an era in which we are frighteningly exposed to potential cataclysm, enhanced by our own seeming ability to warm the earth and increase the probability of extreme climatic events." What lesson has he gleaned from his

historical review? "Like many civilizations before us, we've simply traded up in scale, accepting vulnerability to the big, rare disaster in exchange for a better ability to handle the smaller, more common stresses such as short-term droughts and exceptionally rainy years." He worries that, terrible as the death tolls have been in recent years from famines and natural disasters, the potential demographic tragedies from the inevitable climatic swings of the future could be simply horrific — especially as governments around the

world do not seem to be paying enough attention to this probability.

Whether or not one agrees with Fagan's conclusions, his arguments are clearly drawn and deserve careful scrutiny. *The Long Summer* is a compelling and fascinating book that should interest a broad scholarly audience and general readers alike. ■

Jeremy A. Sabloff is at the University of Pennsylvania Museum of Archaeology and Anthropology, 3260 South Street, Philadelphia, Pennsylvania 19104-6324, USA.

Together forever?

One of Us: Conjoined Twins and the Future of Normal

by Alice Dormurat Dreger

Harvard University Press: 2004. 224 pp.

\$22.95, £14.95

Jonathan Cole

Chang and Eng Bunker, the original (and self-styled) Siamese twins, travelled widely, married and had 22 children between them. As far as we know, they seemed happy with their conjoined state and only contemplated separation, late in their lives, to please their wives. A pair of American conjoined twins, Abigail and Brittany Hensel, who share one body, have required little medical care and live happily, enjoying swimming and cycling. And Yvonne and Yvette McCarther did not regard themselves as handicapped or deformed but merely different.

In *One of Us*, medical historian Alice Dormurat Dreger uses the personal accounts of such twins, and those of their parents and medical carers, to invite us to see conjoined twins as they see themselves. Many enjoy well-adjusted, rich lives, made possible by the development of cooperation strategies; these often work so well that Dreger suggests we could all learn from them. They find ways of becoming individuals, with different interests, while retaining a deep bond with their twin. As Lori Schappell says, "I'm a conjoined twin...but I do not live a conjoined life. The only time I think of it is when I'm interviewed. It's just an integral part of my life." Dreger suggests that although being conjoined is not preferable to being a singleton, conjoined twins generally accept it as part of their identity. The Iranian conjoined twins Ladan and Laleh Bijani were apparently the first in history to ask for a separation; most seem not to have contemplated it.

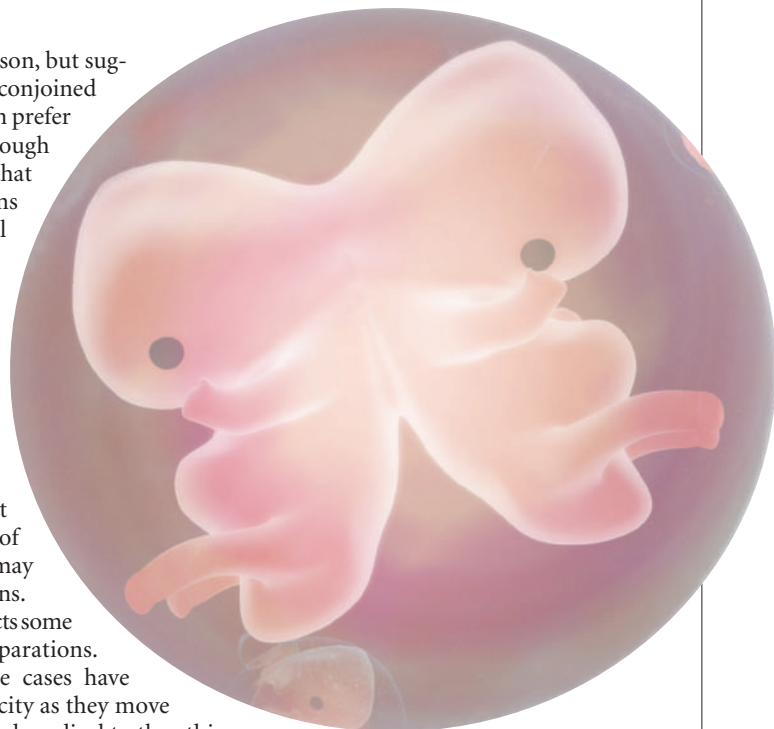
The types of conjoining vary, from the Bunkers' sharing of some abdominal organs to the joining of heads or pelvises. These differences are crucial when separation is contemplated. Dreger discusses whether surgery is done for medical reasons or for psychosocial reasons, to 'normalize' appearance. She admits that life can be easier for a singleton

than a conjoined person, but suggests that many conjoined twins accept and even prefer being conjoined, although there is evidence that some conjoined twins have been grateful for separation. They may also be more vulnerable, as such operations are typically performed in childhood, before their consent can be obtained. Generalization is difficult, however, not least because, like the rest of us, conjoined twins may have differing opinions.

Dreger deconstructs some of the more recent separations. Although rare, these cases have attracted huge publicity as they move from the personal and medical to the ethical, religious, social and legal. Balancing long-term survival of one or both conjoined twins with their associated morbidity must be extraordinarily difficult. In discussing these high-profile cases, Dreger draws out the presumption that she finds underlying most singletons' opinions, whether medical or legal, that life as a conjoined twin is somehow less desirable. She asks us to compare saving one conjoined twin by accelerating the death of the other with asking one of two singleton twins to make a similar sacrifice.

In this, and throughout the book, runs the thread implicit in her subtitle, "the future of normal". She questions whether difference has to be viewed as an impairment and whether impairment is tragic. The mother of conjoined twins is quoted as saying that "the only tragedy is in their interpretation of the girls' situation, since Ruthie and Verena are happy kids". Disability arises not from the impairment but from the response to it in those around, and so is socially induced; it is "about a failure to build ramps [rather] than legs that don't move".

One consequence of viewing difference as an impairment is the pressure for medical



intervention and normalization. The arguments that Dreger gives against these arose when the disabled borrowed from other rights movements to challenge restrictions imposed by difference, whether in anatomy, employment or gender orientation. Why, Dreger asks, "should people with unusual anatomies be treated as if their socially challenging bodies are inherently diseased?" Rather than guaranteeing a child a 'normal' body, we can "try to guarantee a just world".

These are huge questions, not least in considering the situation of any individual with impairment in the world as it is now, which guide the approach of medical workers. To those coming to these arguments afresh, they may prove quite a stretch. Dreger makes no claim to know all the answers but, by taking their side so eloquently, she invites us to see conjoined twins as "no more broken than the rest of us". This book is an eloquent and humane plea to see conjoined twins, and others with impairment and disability, as 'us' and not 'them'. ■

Jonathan Cole is at the universities of Bournemouth and Southampton, and is a clinical neurophysiologist at Poole Hospital, Longfleet Road, Poole BH15 2JB, UK.

Insignificance

Dark matter and dark energy: they might be more abundant than the stuff we are made of, but are they any more interesting?

Sean Carroll

Humans seem to be extremely unimportant in the grand scheme of the Universe. This insight is often associated with Copernicus, who suggested (although not for the first time) that the Earth was not the centre of the Solar System. A bigger step towards calibrating our insignificance was taken by Edwin Hubble, who determined that astrophysical nebulae are really separate galaxies in their own right. We now think there are about one hundred billion such galaxies in the observable Universe, with perhaps one hundred billion stars per galaxy.

But a metaphysically distinct blow to our importance came with the introduction of the idea of dark matter — we are not even made of the same stuff that comprises most of the Universe. The need for dark matter, in the sense of ‘matter we cannot see’, was noticed in 1933 by Fritz Zwicky, when studying the dynamics of the Coma cluster of galaxies. When galaxies are orbiting each other, their typical velocities will depend on the total mass involved, but when we observe clusters of galaxies, the velocities are consistently much higher than we would expect from the mass we actually see in stars and gas. Vera Rubin and others have driven the point home by examining individual galaxies. As we move away from the central galactic region, the velocity of orbiting gas becomes systematically higher than it should be. These observations imply the existence of an extended, massive halo of dark matter. Indeed, the picturesque galaxies we see in astronomical images are really just splashes of visible matter collected at the bottom of these more substantial, yet invisible, halos.

Of course, the air we breathe is invisible and transparent, just like dark matter. A sensible first guess might be that the extra mass we infer is ordinary matter, just in some form we cannot see. But we have independent ways to measure the amount of ordinary matter, through its influence on the early-Universe processes of primordial nucleosynthesis and the evolution of density perturbations. These constraints imply that ordinary matter falls far short of what is needed to explain galaxies and clusters (perhaps one-fifth of the total). Not only is dark matter ‘dark’, it is a completely new kind of particle — something outside the standard model of particle physics, something not yet detected in any laboratory here on Earth.



In *Insignificance* (1985), the surreal meeting of Einstein and Monroe explores relativity and our place in the Universe.

And we have not even mentioned dark energy — the mysterious form of energy that is smoothly distributed throughout space and (at least approximately) constant through time. Independent observations of high-redshift supernovae, the microwave background radiation and the distribution of large-scale structure all require the existence of dark energy. The featureless, persistent nature of dark energy convinces us that it is not even a particle at all. About 70% of our current Universe is dark energy and 25% is dark matter. This leaves all the stuff we have directly observed at a paltry 5% of the whole Universe.

We infer the existence of dark matter and dark energy indirectly, through the influence of their gravitational fields on the ordinary matter that we can see. A famous example of such an inference is the discovery of Neptune, whose existence had been postulated to explain the motion of Uranus. But there is a famous failure, as well — an inner planet ‘Vulcan’ was once postulated to explain the discrepant motion of Mercury. The correct explanation is that our understanding of gravity was flawed — replacing Newtonian gravity with Einstein’s general theory of relativity accounts perfectly for Mercury’s orbit. Could a modification of Einstein’s theory account for dark matter or dark energy? Perhaps, but constructing such a theory has resisted the attempts of even the most ambitious theorists.

A minimal model of dark matter and dark energy — no noticeable interactions between each other or with ordinary matter — fits the data very well. But if 95% of the Universe is in the form of unseen substances, does this not mean that there is the possibility of hidden structure? Might the dark sector be a fascinating place, with its own intricate interactions — perhaps even a kind of intelligent life? Is there a ‘dark light’ that we do not see, radiating and absorbing in the dark Universe?

Probably not. If dark matter particles interacted in a way analogous to ordinary matter, they would also behave analogously. In particular, they would collide and cool, settling into the central regions of galaxies, rather than dispersing into extended halos. If the dark energy were coupled to matter of any sort, it would tend to give rise to unseen forces. Although

it is hard to absolutely rule out any kind of interesting physics in the dark sector, it is an impressive fact that the minimal model fits a broad variety of phenomena with high precision. More often than not, adding new interactions makes the fit worse rather than better.

Still, it is worth keeping an open mind. The minimal model has some apparent discrepancies, for example, substructure in galaxies and the concentration of dark matter near galactic centres seem to be less than we would predict. Perhaps this is because of new interactions. More likely, our current ability to make accurate predictions is not very good in these regimes.

Sometimes ‘most’ does not imply ‘most interesting’. If human pride is wounded by the revelation of our secondary status in the cosmic inventory, we can take some solace in the recognition that the Universe of dark matter is a cold, quiet place indeed. ■

Sean Carroll is in the Department of Physics, University of Chicago, 5640 S. Ellis Avenue, Chicago, Illinois 60637, USA.

FURTHER READING

Krauss, L. *Quintessence: The Mystery of the Missing Mass* (Basic Books, New York, 2001).

Peebles, P. J. E. *From Precision Cosmology to Accurate Cosmology* online at <http://arxiv.org/abs/astro-ph/0208037>.

Rees, M. *Our Cosmic Habitat* (Princeton Univ. Press, Princeton, 2003).

Dust-filled doughnuts in space

Julian Krolik

The first images of an extragalactic object to have been captured using infrared interferometry reveal the doughnut-shaped cloud of dust that obscures the heart of a nearby active galaxy.

Active galactic nuclei are among the most exotic objects in the Universe. They radiate as much light as an entire galaxy from a region the size of the Solar System and, unlike stars, spread that light over the entire electromagnetic spectrum, from radio waves to γ -rays. Thanks to observational advances in the past decade, we now have good evidence that the 'engine' powering each of these objects is a black hole, weighing anywhere from millions to billions of times as much as the Sun. Fortunately for the galaxies that house them (but unfortunately for distant observers like us), active galactic nuclei are often shrouded in opaque gas and dust that block a view of them from most directions. The light from them, intercepted by these dust clouds, is degraded to infrared light that tells us little about the fascinating activity deep inside.

To the further frustration of astronomers, the dust clouds are so close to the active nucleus that their structure cannot easily be made out: rough estimates put their typical size on the sky at 0.01 arcseconds (a few hundred-millionths of a degree), and even the Hubble Space Telescope can't resolve anything smaller than about 0.1 arcseconds. But this barrier is at last being breached. On page 47 of this issue, Jaffe *et al.*¹ present infrared images of an active nucleus that have a resolution of 0.01 arcseconds, achieved through interferometry — the careful combination of images from different telescopes placed some distance apart. The obscuring dust clouds, a scant several light years from a supermassive black hole, have now come into view.

Light can be thought of either as a collection of individual energy packets called photons or as an electromagnetic wave. Interferometry exploits the wave aspect of light in that it hinges on comparing the phases of electromagnetic waves from a single source when they strike different telescopes. Measuring and retaining this phase information is a formidable technical challenge, so interferometry is not a discipline for the faint-hearted. Although the technique has been in regular use for decades at radio wavelengths, the difficulties increase as the wavelength of the light shrinks. The advance into the infrared region of the spectrum has been accomplished only recently² and, as is usually the case, the first observations were only of nearby, comparatively bright stars^{3–5}.

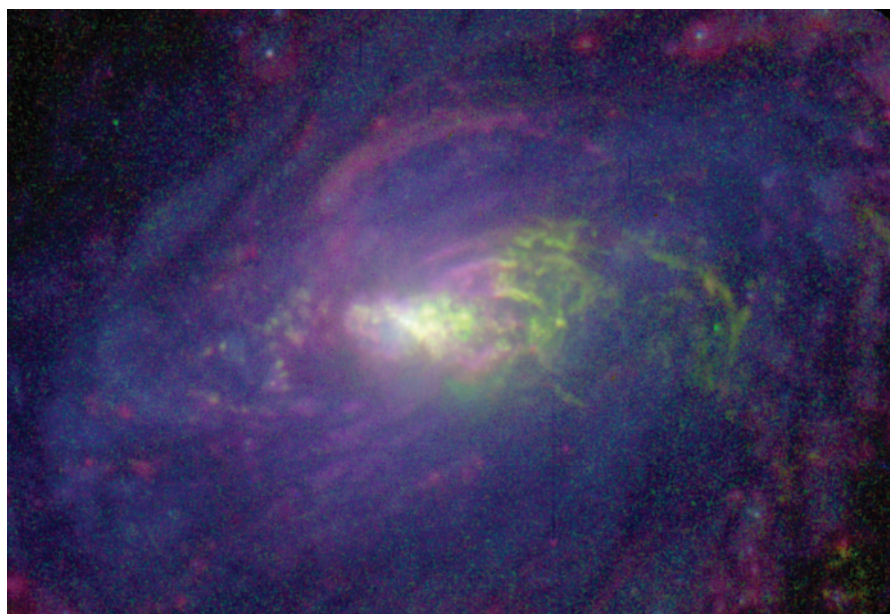


Figure 1 The active galaxy NGC 1068. At its centre is a supermassive black hole. Using infrared interferometry, Jaffe *et al.*¹ have created a high-resolution image that reveals a warm torus, or doughnut shape, of dust enshrouding the black hole. The three (false) colours in this image correspond to different optical wavelengths. (Image courtesy of Cherie Miskey and Fred Bruhweiler, Catholic Univ. America.)

Jaffe *et al.*¹ have used the interferometer formed by the European Southern Observatory's Very Large Telescope and several smaller telescopes on the same Chilean mountain⁶. With this apparatus, high-resolution infrared images of faint, distant objects have been obtained for the first time. Indeed, the data shown in Jaffe and colleagues' paper¹ are the first to be derived from infrared interferometry on anything outside the Milky Way. And what Jaffe *et al.* see is as interesting as the gear they built to see it with. Their images show a geometrically thick ring of dust that is extremely close to the central black hole in the nearest powerful active galaxy, NGC 1068 (Fig. 1). The warmest dust is no more than two or three light years from the very centre of the beast.

This result is a dramatic confirmation of inferences made many years ago, but which were not entirely accepted because they were so hard to understand. Since the late 1980s, many have believed that the dust clouds surrounding active galactic nuclei are not higgledy-piggledy, but are instead arranged more or less like a thick doughnut. As a result, observers whose sight-lines fall near to the equatorial plane of any particular active

galactic nucleus (most of the Universe, in fact) are relegated to seats with an obscured view of its centre; the favoured minority close to the axis of the doughnut get to see the full show unobstructed.

But this picture was difficult to accept because, in a gas cool enough to permit dust to survive, the thermal motions would be too slow to keep the doughnut puffed up and thick; the gravity of the central black hole would ensure that the doughnut collapsed to look more like a CD, if the only resistance to it were random thermal motions⁷. Many speculative suggestions have been made to account for the geometrical thickness of the dust doughnut^{7–10} but none has gained general credence. Partly to avoid this dynamical conundrum, alternative geometries for the dust (for example, a thin but warped disk^{11,12}) have been suggested, but these are problematic for other reasons⁷. Thanks to these new observations¹, we now know that, at least in the nearest and brightest example of NGC 1068, a thick doughnut is in fact the right shape.

The fact that most active galactic nuclei cannot be seen clearly has many implications. Most obviously, it means that it is hard

to find them, so censuses of massive black holes probably underestimate their numbers by as much as a factor of five. If most active nuclei are visible only in difficult-to-observe infrared light, their total power may actually be so large as to rival the total output of all the stars in the Universe¹³.

After languishing for a decade largely through lack of data, this field should now see a revival, as it is refreshed by detailed infrared imaging. The dynamical problems guessed at years ago can be brought into clearer focus; with any luck, direct images of these structures may yield clues to their solution. ■

Julian Krolik, of Johns Hopkins University, is currently at the Institute of Astronomy,

Madingley Road, Cambridge CB3 0HA, UK.

e-mail: jhk@ast.cam.ac.uk

1. Jaffe, W. *et al.* *Nature* **429**, 47–49 (2004).
2. Townes, C. *et al.* *Proc. SPIE* **3350**, 908–932 (1998).
3. Bester, M. *et al.* *Astrophys. J.* **463**, 336–343 (1996).
4. Lopez, B. *et al.* *Astrophys. J.* **488**, 807–826 (1997).
5. Malbet, F. *et al.* *Astrophys. J.* **507**, L149–L152 (1998).
6. Leinert, C. *et al.* *Proc. Soc. Photo Optical Instrum. Eng.* **4838**, 893–904 (2003).
7. Krolik, J. H. & Begelman, M. C. *Astrophys. J.* **329**, 702–711 (1988).
8. Pier, E. A. & Krolik, J. H. *Astrophys. J.* **399**, L23–L26 (1992).
9. Yi, I., Field, G. B. & Blackman, E. G. *Astrophys. J.* **432**, L31–L34 (1994).
10. Konigl, A. & Kartje, J. F. *Astrophys. J.* **434**, 446–467 (1994).
11. Sanders, D. B., Phinney, E. S., Neugebauer, G., Soifer, B. T. & Matthews, K. *Astrophys. J.* **347**, 29–51 (1989).
12. Pringle, J. E. *Mon. Not. R. Astron. Soc.* **292**, 136–147 (1997).
13. Fabian, A. C. & Iwasawa, K. *Mon. Not. R. Astron. Soc.* **303**, L34–L36 (1999).

Regenerative medicine

Self-help for insulin cells

Ken Zaret

Insulin-producing β -cells in the adult pancreas were thought to derive from pancreatic stem cells. But it seems that they arise abundantly from β -cells themselves, offering a new outlook on regenerative medicine.

After we eat a meal, we must store excess sugars so that energy reserves will be available later on. Excess glucose is sensed by β -cells in the pancreas, which respond by secreting the hormone insulin into the bloodstream. Insulin, in turn, causes various cells in the body to store glucose. In people with type I diabetes, the immune system destroys β -cells, resulting in a lifelong dependency on insulin treatments. Recently, marked advances have been made in transplanting clusters, or 'islets', of β -cells from human cadavers into type I diabetics, making insulin treatments unnecessary¹. But that source of β -cells is limited, so the science of producing new β -cells has become a red-hot topic, sparking a flurry of studies into how the adult body itself generates insulin-producing cells. On page 41 of this issue, Dor and colleagues² describe how they used a genetic lineage-marking approach to shed light on the origin of β -cells in adult mice. Their finding that these cells arise mostly from pre-existing β -cells, and not from pancreatic-duct or stem cells, counters prevailing hypotheses in the field.

Analysis of human pancreatic tissue has shown that insulin-producing cells can first appear in and around the walls of the pancreatic ducts, in addition to the β -cell islets³. For this reason, and because of their high proliferative rates in experimental models of pancreas regeneration⁴, pancreatic-duct cells have long been thought to be a source of β -cells. Various stem-cell types have also been correlated with the appearance of new β -cells (discussed in ref. 2). This fits with the situation in the skin and intestine: cell-lineage studies have shown that unspecialized

stem cells generate the specialized cells of these tissues.

Correlative studies, however, assess cell position, expressed genes and proliferative states, which can change so rapidly that they do not definitively track cell fates. The pancreatic field of research has been misled by this before: the appearance of early embryonic cells expressing both insulin and glucagon suggested that they later give rise to cells that express only one or the other hormone. But a genetic method to mark cell lineages showed that mature insulin-secreting β -cells and glucagon-producing cells derive from a different embryonic precursor⁵.

Dor and colleagues² have now used this genetic method to indelibly mark the DNA of adult mouse β -cells, so that they could definitively track the cells and their descendants during normal and stimulated cell turnover. The authors inserted into mice a hybrid gene containing the regulatory DNA sequence (the promoter) of the insulin gene, linked to a gene for a recombinase — an enzyme that can rearrange specific DNA sequences in a cell (Fig. 1a). The insulin regulatory sequence causes the recombinase to be expressed only in β -cells, not in duct cells or other cell types. An additional twist is that the recombinase is inactive unless the animal is treated with a synthetic hormone.

When Dor *et al.* administered a pulse of this hormone to the mice, it activated the recombinase, which produced a small genetic rearrangement, or mark, in the cells. Just after the hormone pulse, all of the marked cells examined were insulin-positive β -cells. The authors then tracked descendants of the marked cells for up to a year and determined

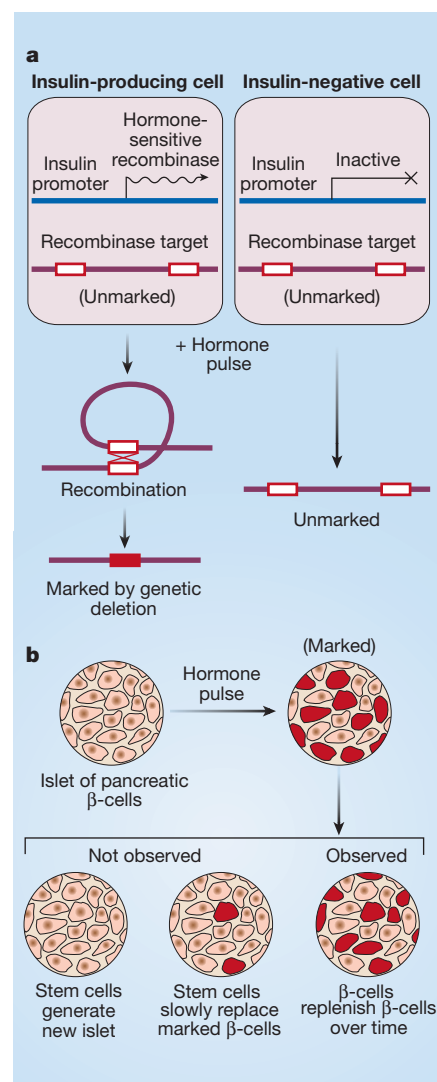


Figure 1 Where do adult β -cells come from? Dor and colleagues² used the technique shown to investigate this question. **a**, Genetic marking. The authors inserted into mice a gene that has a regulatory region (promoter) from the insulin gene, fused to a gene encoding a recombinase enzyme. The hybrid gene is active only in insulin-producing cells — β -cells — and its recombinase product is active only in the presence of a synthetic hormone. When the hormone is added, the recombinase is activated in β -cells and causes two other genetic regions to 'recombine', producing a deletion that acts as a specific marker of β -cells. **b**, Cell-lineage analysis. In response to the hormone, many cells in a cluster (islet) of β -cells become marked; the authors then monitored what happens during normal cell turnover or following pancreatic damage. If stem cells (which are unmarked) were to generate new β -cells, then unmarked β -cells would take over the islet; if they were to generate entirely new islets, those islets would contain unmarked β -cells. But a sizeable proportion of the β -cells remained marked during cell turnover, indicating that they were produced from the original marked β -cells.

their fate, during both normal cell turnover and after part of the pancreas was removed to stimulate regeneration.

Surprisingly, the authors found that the percentages of marked β -cells, during both normal turnover and pancreatic regeneration, remained stable, and were not diminished by β -cells arising from insulin-negative progenitors (duct or stem cells; Fig. 1b). In other words, many of the β -cells generated over the study period were marked cells, indicating that they originated from cells that expressed the recombinase at the time the hormone pulse was given — that is, β -cells. The authors' quantification showed that the contribution of non- β -cells — duct or stem cells — was minimal. They also failed to find new, unmarked islets forming during normal β -cell turnover; such unmarked islets would be predicted if other cell types could generate β -cells.

What questions arise from this work? First, the technique that Dor and colleagues used to stimulate pancreas regeneration — partial pancreatic removal — causes acute damage and cellular responses⁴. Their findings are reminiscent of the way in which liver regenerates after partial removal^{6,7}. But what happens in models of long-term pancreatic damage⁸? Another question concerning the experiments themselves is whether the recombinase gene used is a little 'leaky', producing some marked non- β -cells that escaped detection and gave rise to marked β -cells. But even if so, these non- β -cells can have given rise only to β -cells: other pancreatic cell types were not marked during regeneration.

With regard to applications of the work, Dor and colleagues' findings raise the prospect of removing β -cells from adult human cadavers, stimulating cell replication, and transplanting the resulting large quantities of β -cells into diabetic patients. This is still a long way off, but the authors' results indicate that research should now focus on the fundamentals of β -cell replication. One question to be addressed is whether human β -cells replicate like mouse β -cells. Human β -cells are known to be able to proliferate, albeit less so than duct cells⁹, but the extent to which they can generate new β -cells is unknown. Also, we need to know more about how β -cells interact with other cells in and around the islet, such as blood-vessel cells^{10,11}.

Another potential source of β -cells for transplantation is human embryonic stem (ES) cells¹². But if we can generate β -cells from adult β -cells, why contend with ES cells and their associated ethical implications? First, there are insufficient cadaveric β -cell donors. More significantly, ES-derived β -cells are generally 'younger' than cells from cadavers, and so less likely to have accumulated cellular and chromosomal aberrations. Consider, for instance, transplanting β -cells from a 50-year-old cadaver into a diabetic child, and

hoping that the cells will remain glucose-responsive and not become carcinogenic over a second human lifespan. Another reason for continuing ES-cell research is that genetic manipulation¹³ or therapeutic cloning¹⁴ might eventually be used to ameliorate the recipient's immune intolerance to β -cells.

Nonetheless, given the intense activity in the field of β -cell regeneration, the current focus on adult stem cells, and the early state of ES-cell research, the work of Dor and colleagues² can be considered a paradigm shift. It doesn't exclude the possibility that, in some diseases, adult stem cells contribute to the β -cell population. But it does shine light on a resource for insulin-producing cells that has been there all along: the β -cell itself. ■

Ken Zaret is in the Cell and Developmental Biology Program, Fox Chase Cancer Center, 333 Cottman

Avenue, Philadelphia, Pennsylvania 19111, USA.

e-mail: zaret@fccc.edu

1. Shapiro, A. M. N. *Engl. J. Med.* **343**, 230–238 (2000).
2. Dor, Y., Brown, J., Martinez, O. I. & Melton, D. A. *Nature* **429**, 41–46 (2004).
3. Butler, A. E. *et al.* *Diabetes* **52**, 102–110 (2003).
4. Bonner-Weir, S., Baxter, L. A., Schupp, G. T. & Smith, F. E. *Diabetes* **42**, 1715–1720 (1993).
5. Herrera, P. L. *Development* **127**, 2317–2322 (2000).
6. Rhim, J. A., Sandgren, E. P., Degen, J. L., Palmiter, R. D. & Brinster, R. L. *Science* **263**, 1149–1152 (1994).
7. Overturf, K., al-Dhalimi, M., Ou, C. N., Finegold, M. & Grompe, M. *Am. J. Pathol.* **151**, 1273–1280 (1997).
8. Wang, R. N., Kloppel, G. & Bouwens, L. *Diabetologia* **38**, 1405–1411 (1995).
9. Tyrberg, B., Ustinov, J., Otonkoski, T. & Andersson, A. *Diabetes* **50**, 301–307 (2001).
10. Lammert, E., Cleaver, O. & Melton, D. *Science* **294**, 564–567 (2001).
11. Yoshitomi, H. & Zaret, K. S. *Development* **131**, 807–817 (2004).
12. Thomson, J. A. *et al.* *Science* **282**, 1145–1147 (1998).
13. Chamberlain, J. R. *et al.* *Science* **303**, 1198–1201 (2004).
14. Hwang, W. S. *et al.* *Science* **303**, 1669–1674 (2004).

Biomechanics

Fast fish

Adam P. Summers

Swift-swimming, open-ocean hunters such as mako sharks and tunas need a big engine. Despite their long separation in evolutionary terms, the internal drive systems adopted by these fishes are much the same.

To an underwater observer, mako sharks look a lot like tunas. There is an added *frisson* from the dental battery of the shark, but the two fishes share a common body plan, colour pattern and even swimming style. On page 61 of this issue, Donley and her colleagues¹ demonstrate that, after 400 million years of separate evolutionary trajectories, these two high-speed predators have converged on solutions to the problem of swimming fast that go from skin to skeleton.

Tunas and mackerels, as well as makos and their relatives, appear in the fossil record

about 60 million years ago. They roamed the warm Tethys Sea, hunting the newly diversifying radiations of other fishes. Their common ancestor dates back to the Carboniferous, when the bony and cartilaginous fishes diverged².

The 'thunniform' body plan is distinctive and has evolved independently several times. The finely streamlined, teardrop-shaped body, with a high-aspect-ratio tail fin and a narrow tail region, is familiar from several speedy denizens of the deep — including dolphins, marlin and makos, and even the extinct ichthyosaurs. Tunas and dolphins also

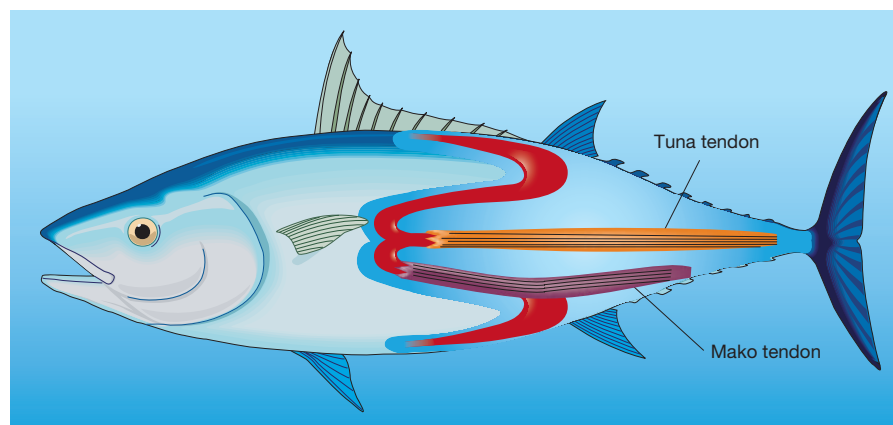


Figure 1 Marine power transmission. The muscle that propels a fish has a complex three-dimensional structure. The boundaries of each block of muscle are defined by a sheet of connective tissue, which is bent into a zig-zag with two cones of muscle pointing forwards and two pointing backwards. In swift-swimming tunas, power is transmitted by a prominent tendon that arises between the two forward-pointing cones. Donley *et al.*¹ show that a tendon has arisen convergently in the mako shark from the backward-pointing cone of connective tissue.

share a swimming style; they have lost the whole-body fishy wriggle in favour of a businesslike sweep of the tail that leaves the predatory front end of the body relatively steady. Donley *et al.* show that the mako shark also has this swimming style. This lends support to the hypothesis that thunniform ichthyosaurs swam swiftly, generating thrust with their tail rather than their whole body³.

Fishes typically have two distinct muscle types on their flanks: a large mass of white muscle and a smaller band of red muscle running the length of the animal just under the skin. Fishes cruise along by throwing their body into waves with alternate contractions of the aerobic, fatigue-resistant red muscle. For short bursts the larger, anaerobic and rapidly fatiguing white muscle is activated.

In both tunas and makos the body muscle mass is shifted well forward, giving them a distinctive broad-shouldered look when viewed from above. The red muscle is also located within the mass of white muscle, nearer the vertebral column, rather than close to the skin. This anatomy allows heat from the continuously activated red muscle to be retained in the core of the fish, leading to local warm-bloodedness.

Shifting the red muscle forward and towards the midline requires modifications of the force-transmission system that drives the tail. As anyone who has picked at a plaice fillet can tell, fish body musculature is divided into a series of blocks with a complex three-dimensional shape. Between each block is a 'myoseptum', a sheet of densely collagenous connective tissue that acts as a sheet of tendon⁴.

In most fishes these blocks form a pair of W's, lain on their side and arranged one above the other. In tunas and mackerel the points of the W's are extended into very long tendons that run much of the length of the body and insert on the tail (Fig. 1). The red muscle, acting on these long tendons, transmits its force over a far greater distance than in other fishes⁵.

The anatomy of the myosepta of makos is quite different, with a linearly arrayed tendon arising out of one of the legs of the W. This tendon is both distinct and quite long, in some cases reaching 19% of the total length of the animal (Fig. 1). Towards the tail of the fish, these myoseptal tendons converge into a heavy collar of connective tissue. Donley *et al.*¹ show that, despite the difference in shape, mako tendons play an identical role in shifting the force generated by the red muscle mass towards the back end of the fish.

In Donley and colleagues' experiments, mako sharks were allowed to swim in the aquatic version of a treadmill: a broad-bore tube that recirculates water from the tail of the fish to the head, allowing the animal to swim in place. Using sonomicrometry, a technique that involves implanting tiny piezoelectric crystals in the muscle, they

measured the shortening of the white and the red muscle fibres.

During passive undulation, when neither muscle type is active, the red and white muscle changed length in perfect lock step. When the mako started swimming actively, requiring activation of the red cruising muscle, there was a clear shift in the timing of shortening. The red muscle shortens well after the white, a shift also seen in tunas. This change in timing means that the red muscle is acting further along the body, causing curvature more towards the tail, while the white muscle causes local curvature changes.

For decades, the similarities between mako sharks and tunas have been the subject of speculation. After all, understanding the mechanisms behind their locomotion could

lead to high-speed autonomous underwater vehicles. The data from these two high-speed swimmers seem clearly to endorse a solution that puts as much emphasis on the placement of actuators as on the overall shape of the vehicle. ■

Adam P. Summers is in the Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, California 92697-2525, USA.

e-mail: asummers@uci.edu

1. Donley, J. M., Sepulveda, C. A., Konstantinidis, P., Gemballa, S. & Shadwick, R. E. *Nature* **429**, 61–65 (2004).
2. Bernal, D., Dickson, K. A., Shadwick, R. E. & Graham, J. B. *Comp. Biochem. Physiol.* **129**, 695–726 (2001).
3. Motani, R. *Nature* **415**, 309–312 (2002).
4. Summers, A. P. & Koob, T. J. *Comp. Biochem. Physiol.* **133**, 1159–1170 (2002).
5. Westneat, M. W., Hoese, W., Pell, C. A. & Wainwright, S. A. *J. Morphol.* **217**, 183–204 (1993).

Earth science

Hot metal

Carl B. Agee

The solubility of oxygen in molten iron increases at high temperature. Could this explain why Earth's mantle is poor in iron oxide, whereas the mantle of Mars, which formed under cooler conditions, is not?

The fate of light elements, such as oxygen, sulphur, carbon and hydrogen, during the formation of planetary cores has been hotly debated for decades. Forty years ago, Birch¹ estimated that about 10% of Earth's core must be made of one or more of the light elements, to account for the density deficit seen in seismic observations when compared with a hypothetical core of pure iron metal. The advent of high-pressure, high-temperature technology in the 1980s has enabled the behaviour and distribution of elements between the mantle and core to be examined under conditions that simulate those in the deep planetary interior. Results from some of the first exploratory experiments² suggested that sufficient oxygen would dissolve in iron metal under the extreme pressures at Earth's core–mantle boundary to account for the density deficit.

But Rubie *et al.*³ have evaluated data on oxygen partitioning between mantle and core materials — from experiments performed by them and others, under extreme conditions^{3–5} — and conclude that higher temperatures dramatically increase the amount of oxygen that will dissolve in molten iron metal (page 58 of this issue). In contrast to earlier work, Rubie *et al.* find that pressure has the opposite effect, decreasing the affinity of oxygen for molten iron metal. But if temperature and pressure have opposite effects on oxygen solubility in molten metal, which, if either, wins out in the high-pressure, high-temperature environment of planetary interiors? The deciding factors, it seems, are the size of the planet and

the depth of its hypothesized 'magma ocean'.

About four billion years ago, the newly formed Earth and its neighbouring planets were subjected to an intense meteor bombardment — so intense that the collisions might have melted a considerable depth of the fledgling planets' rocky surfaces into oceans of magma. Rubie *et al.*³ note that if a magma ocean on Earth had occupied a slightly greater volume than that of the modern upper mantle, extending to less than 800 km below the planet's surface, then the magma could only have been as hot as 2,500 K — not hot enough to allow the temperature effect to dominate oxygen solubility. They calculate that temperature would have made a difference only if the magma ocean extended well into the lower mantle, perhaps as deep as 1,800 km, where temperatures need to be 3,500 K or above for rock to melt (Fig. 1, overleaf). This depth of magma ocean is broadly consistent with the trace levels of the siderophile, or 'iron-loving', elements that remain in Earth's mantle today. These elements have been used as geobarometers or thermometers to estimate that the magma ocean was as deep as 1,200 km, with temperatures as high as 3,500 K and pressures of 50 gigapascals at its base^{6,7}.

How did this hot, deep magma ocean process and transform the primordial material from which the early Earth formed? Imagine building the Earth with a mix of silicates and metal in the proportions found in meteorites; about half of the iron would exist in its metal form and the other half as iron oxide, mostly as a constituent of silicate

NATURE

100 YEARS AGO

In a paper read before the American Philosophical Society Mr. Percival Lowell discusses the 375 drawings of the Martian surface made by him during the opposition of 1903. Having plotted the values allotted to the "visibility" of eighty-five canals, at different periods, with regard to the time of their minima visibilities after the Martian summer solstice, he found that these minima appeared in regular sequence from the North Pole towards the equator. Mr. Lowell believes that the canals are strips of vegetation dependent for their growth — and therefore for their visibility — upon the simultaneous presence of sunlight and water, and he points out that on a planet, such as the earth, where water is constantly present all over the surface, the appearance of vegetation solely depends upon the amount of sunlight received; therefore in the northern hemisphere it simply progresses northward with the sun. On the other hand, he concludes, from his curves, that there is no constant supply of moisture on the surface of Mars, and, therefore, although the sun may have reached the summer solstice, it is not until the snowcap melts and looses the water supply that vegetation appears.

From *Nature* 5 May 1904.

50 YEARS AGO

Prof. E. Schrodinger's recent book discussing the relevance of ancient Greek thought to present-day scientific philosophy touches on very many of the problems which have been debated throughout the whole history of Western civilization. He closes it... with the thought... that "for the purpose of constructing the picture of the external world, we have used the greatly simplifying device of cutting our own personality out, removing it...". But has not recent biological thought added to the picture a factor which Schrodinger fails to take into account? We believe now that the intellectual-sensory apparatus by which we make contact with the external world arose through a process of evolution... If this is true, it is no longer adequate to frame a discussion in terms of a dichotomy between "our personality" and an impersonal world which we arrive at by "cutting our personality out". The faculties by which we arrive at a world view have been selected so as to be, at least, efficient in dealing with other existents.

From *Nature* 8 May 1954.

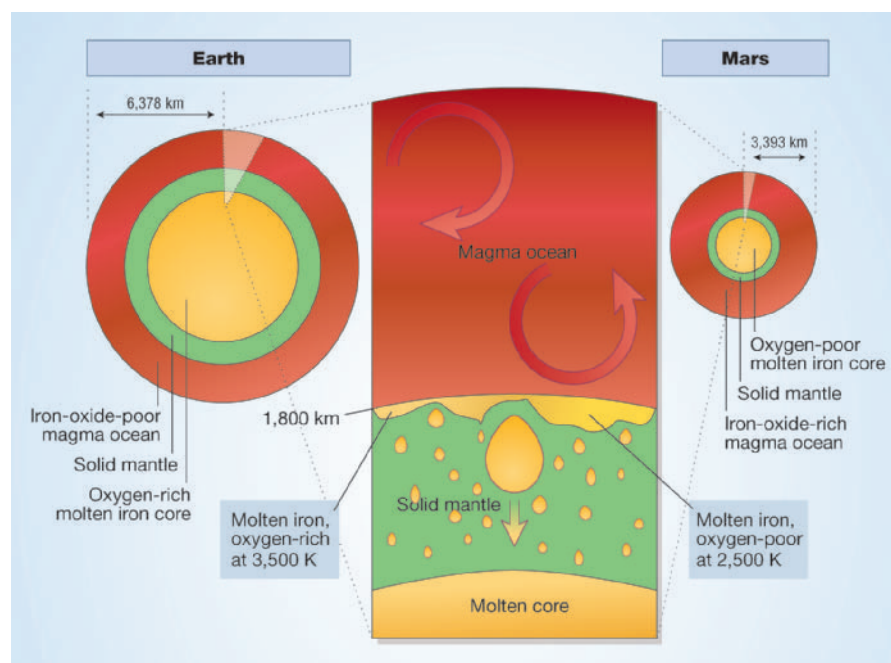


Figure 1 Magma oceans on early Earth and Mars. Bombardment by meteors billions of years ago might have melted the surface of both planets into a magma ocean. On the basis of experiments performed under high-pressure and high-temperature conditions, Rubie *et al.*³ suggest that Earth's magma ocean might have extended to a depth of 1,800 km. At the bottom of this ocean, the high temperature and pressure would have caused more oxygen to dissolve in molten iron metal, eventually to be carried down into the planetary core. The magma ocean would have been left depleted of iron oxide. Mars, however, is smaller than Earth; the pressure at a depth of 1,800 km is lower and the temperature would have been too low to influence oxygen solubility. As a result, its magma ocean remained rich in iron oxide.

minerals such as olivine and pyroxene. In a magma ocean, these materials form two liquids — one silicate and the other iron metal. The liquids separate as the higher-density metal rains in droplets downwards through the lower-density silicate ocean. The metal accumulates at the bottom of the magma ocean and then, if pooled into large masses, sinks further through the solid Earth below, eventually occupying the central core region (Fig. 1).

If the magma ocean is about 1,800 km deep, then, according to Rubie *et al.*³, the temperature effect will cause a significant amount of the molten silicate iron oxide to be converted into molten metallic iron containing dissolved oxygen. Thus, the amount of iron oxide in the silicate mantle would decrease and the amount of iron metal rich in oxygen would increase, eventually being incorporated into the core. The result would be a mantle that is about 8% iron oxide by weight and a metallic iron core that is more massive than would be expected from the proportion of metallic iron in meteorites — precisely what is observed in the modern Earth.

Mars, in contrast, is thought to have a mantle that is richer in iron oxide and a less massive metallic core⁸. Mars is smaller than Earth, having only a tenth of its mass, and its internal pressures are about one-third of those in the Earth at any given depth. For

example, the martian core–mantle boundary lies at a depth of 2,000 km, with a pressure of 25 gigapascals. In the Earth, a pressure of 25 gigapascals corresponds to a depth of only 700 km, close to the boundary between the upper and lower mantle. Rubie *et al.* contend that a magma ocean on Mars, even one as deep as 2,000 km and occupying the entire mantle, would not be hot enough for the temperature effect to perturb the accreted proportions of iron oxide and iron metal (Fig. 1). Thus if, as is thought, Mars formed from materials that contained about 18% iron oxide by weight in silicate minerals, this level would not have been affected by a magma ocean or by core formation, and it should be maintained in the martian mantle today. This is entirely consistent with geochemical models based on martian meteorites⁹ and with Mars' moment of inertia as measured by the Mars Pathfinder mission⁹.

Rubie and colleagues' hypothesis³ is a neat explanation of how terrestrial planets of different sizes could end up with dramatically different core mass fractions and mantle iron-oxide compositions, despite being formed from the same primordial material. Simply put, it is possible because Earth's magma ocean might have been under higher pressure, and thus hotter, during core formation than was the magma ocean on Mars. On the other hand, this elegant solution might not cover all the factors that

determine oxygen and iron distribution in a planetary interior. It is conceivable that Earth and Mars formed from distinctly different geochemical reservoirs in the early Solar System. Perhaps Earth formed under conditions that were more strongly reducing than those that existed for Mars — if oxygen was scarce, no transfer of excess oxygen to the core would be required to explain the low abundance of iron oxide in the mantle today.

Finally, it is worth noting that much of our current knowledge of Mars geochemistry and mantle–core evolution is based on a collection of only 25 meteorites that are known to have originated from Mars. With new data from the Mars rovers, the prospect of a sample-return mission and further seismic measurements of the martian interior,

our understanding of the red planet, and its similarities to Earth, can only improve. ■

Carl B. Agee is at the Institute of Meteoritics, and in the Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, 87131-1126 New Mexico, USA.
e-mail: agee@unm.edu

1. Birch, F. *J. Geophys. Res.* **69**, 4377–4388 (1964).
2. Kato, T. & Ringwood, A. E. *Phys. Chem. Miner.* **16**, 524–538 (1989).
3. Rubie, D. C., Gessmann, C. K. & Frost, D. J. *Nature* **429**, 58–61 (2004).
4. Li, J. & Agee, C. B. *Geophys. Res. Lett.* **28**, 81–84 (2001).
5. O'Neill, H. S. C., Canil, D. & Rubie, D. C. *J. Geophys. Res.* **103**, 12239–12260 (1998).
6. Li, J. & Agee, C. B. *Nature* **381**, 686–689 (1996).
7. Chabot, N. L. & Agee, C. B. *Geochim. Cosmochim. Acta* **67**, 2077–2091 (2003).
8. Dreibus, G. & Wänke, H. *Hans Suess Festschrift* **20**, 367–381 (1985).
9. Folkner, W. M. *et al. Science* **278**, 1749–1751 (1997).

Human genetics

An inflammatory issue

Kevin J. Tracey and H. Shaw Warren

People vary naturally in a protein called caspase-12, and hence in their susceptibility to harmful inflammation. This discovery highlights the balance between the protective and destructive effects of immunity.

Our immune systems are adept at warding off intruders, but often the very strength of the immune response can be a problem. As the immunologist and writer Lewis Thomas¹ put it, “Our arsenals for fighting off bacteria are so powerful, and involve so many different defense mechanisms, that we are in more danger from them than from the invaders”. One manifestation of this danger is severe sepsis — a syndrome that is the third leading cause of death in technologically modern societies, after heart disease and cancer², and which is caused by an overreaction of the immune system to microorganisms or their products. Sepsis is thought to be mediated largely by messenger molecules called cytokines, whose production and release are modulated, at least in part, by enzymes termed caspases.

On page 75 of this issue, Saleh and colleagues³ describe a previously unrecognized variation in one such enzyme, caspase-12, that leads to diminished immune responses and an increased susceptibility to bacterial invasion of the bloodstream and severe sepsis. Although caspases are important in programmed cell death (apoptosis), which occurs in sepsis⁴, this variation appears to affect cytokine production rather than cell death. How do these findings fit into our current theoretical framework for understanding severe sepsis?

In 1884, the German physician Robert Koch developed the four ‘postulates’ that bear his name and which are a means of determining whether a causal relationship exists between a particular microorganism

and a particular disease. A century later, advances in molecular biology led to the discovery that certain molecules in our body provide a modern-day interpretation of the postulates for some of the secondary effects of infection. For instance, tumour-necrosis factor (TNF) — a cytokine that is produced by certain types of white blood cell and whose function is to induce inflammation — proved to be both necessary and sufficient to cause shock and tissue injury in animals^{5,6}.

Other molecules capable of mediating inflammation have been implicated in similar, infection-induced syndromes. And abundant evidence from animals indicates that the lethality of severe bacterial infections results from a secondary, uncontrolled inflammatory reaction. In these examples, bacteria stimulate immune cells, which in turn produce mediators that cause distinct disease characteristics. Blocking the host mediators has cured severe sepsis in animal studies, but the approach has not yet been successfully adapted to humans.

In this context, Saleh and colleagues’ results³ raise interesting issues. The authors’ study was triggered by their discovery that the human caspase-12 gene is polymorphic, which means that its sequence varies in different people. The gene encodes either a short or a long version of caspase-12. Sequence analysis of more than 1,000 genomic samples from people of diverse ethnic backgrounds suggested that the long version is much less common than the shorter form, and is found only in people of African origin. Other species, including mice, have only the long version.

A previous study of mice⁷ found that caspase-12 is involved in apoptosis, and hinted that the enzyme might be important in Alzheimer’s disease. But it seems that the situation is different in humans. Saleh *et al.* found that human blood cells possessing different caspase-12 variants show little difference in their apoptotic response to various stimuli. In contrast, human caspase-12 does appear to alter the production of pro-inflammatory cytokines in response to a bacterial component, lipopolysaccharide. Notably, the longer caspase-12 variant leads to the production of fewer cytokines than does the shorter form. The authors’ preliminary analyses also suggest that patients with Alzheimer’s disease are no more likely to have the long caspase-12 variant than the shorter version. But patients of African descent who have bacterial infections severe enough for bacteria to enter the bloodstream and cause severe sepsis are slightly (yet statistically significantly) more likely to have the long variant.

At first blush, these findings seem paradoxical; patients who produce fewer cytokines (such as people with the long caspase-12) might be expected to suffer a lesser, not a greater, degree of sepsis. How can these findings be reconciled with our previous knowledge of the molecular mechanisms involved? The solution probably lies in the dual nature of cytokines: although they can cause harmful inflammatory damage to tissues, they are primarily an essential part of our defences against microorganisms. It is believed that innate immunity — the body’s first response to intruders — requires an early detection system that depends on cytokine-mediated local inflammation. For example, mice that are deficient in Toll-like receptor 4 (TLR4), which detects lipopolysaccharide and starts a cellular cascade leading to cytokine production, are highly resistant to this bacterial product, but more susceptible to bacterial infection⁸. Likewise, patients who receive anti-TNF therapy to reduce inflammation, as well as individuals who have a defect in TLR4, are at increased risk of infection⁹.

People with the long variant of caspase-12 may similarly be more susceptible to infection, because they produce fewer pro-inflammatory cytokines in response to microbial products. As a consequence, their innate immunity presumably fails to control early bacterial replication, so that bacteria invade the bloodstream in overwhelming numbers, resulting in severe sepsis¹⁰. This fits with the fact that — somewhat ironically — many of the treatments developed for severe sepsis increase the risk of infection when administered long-term. We may need to balance this risk against the potential benefits of treating a short-term excess of an inflammatory mediator.

A broader point to make is that data concerning polymorphisms should be

treated with caution, and that studies should therefore ensure that overall numbers are adequate, populations specifically defined, and appropriate controls included that exactly match the patient population. A difficulty with studies of sepsis is that accepted definitions of the syndrome are very broad — perhaps too broad. This has become evident from the many clinical trials of the past decades, which have generally failed to obtain consistent results from similar studies of patients with severe sepsis. Indeed, the current framework for designing clinical trials in this area seems fundamentally flawed. In such trials, patients with infections caused by different microbes, and affecting different organs, have been lumped together under the same diagnosis of severe sepsis. The situation is almost akin to taking patients who have symptoms in the hip caused by osteomyelitis and patients with similar symptoms that are secondary to osteosarcoma, and entering them in the same clinical trial.

What can be done? There is abundant evidence that numerous molecules, as well as an individual's genetic background, influence both protective innate immunity and the development of severe sepsis. Saleh and colleagues' work³ provides a good example of this. The next step might be to select patients and controls for these clinical trials on the basis of clinical, molecular and genomic data. This would allow disease criteria to be defined that are specific to a single molecular derangement, and to be linked to a treatment specific to that derangement. ■

Kevin J. Tracey is at North Shore–Long Island Jewish Research Institute, 350 Community Drive, Manhasset, New York 11030, USA.
e-mail: kjtracey@sprynet.com

H. Shaw Warren is at Harvard Medical School and Massachusetts General Hospital East, 149 13th Street, Charlestown, Massachusetts 02129, USA.
e-mail: warren@helix.mgh.harvard.edu

1. Thomas, L. N. *Engl. J. Med.* **287**, 553–555 (1972).
2. Angus, D. C. *et al. Crit. Care Med.* **29**, 1303–1310 (2001).
3. Saleh, M. *et al. Nature* **429**, 75–79 (2004).

4. Hotchkiss, R. S. *et al. Nature Immunol.* **1**, 496–501 (2000).
5. Tracey, K. J. *et al. Nature* **330**, 662–664 (1987).
6. Tracey, K. J. *et al. Science* **234**, 470–474 (1986).
7. Nakagawa, T. *et al. Nature* **403**, 98–103 (2000).
8. Poltorak, A. *et al. Science* **282**, 2085–2088 (1998).
9. Agnese, D. M. *et al. J. Infect. Dis.* **186**, 1522–1525 (2002).
10. Munford, R. S. & Pugin, J. *Am. J. Respir. Crit. Care Med.* **163**, 316–321 (2001).

Cell biology

Designer prions

Daniel C. Masison

Prions are clumps of misshapen proteins that can be passed between cells without the need for genetic intermediaries. The parts of the proteins that account for such infectivity are now being dissected.

Alzheimer's disease, type II diabetes and prion diseases — mad cow disease being the most notorious — are all characterized by the accumulation of misshapen proteins into aggregates in various parts of the body. Of these disorders, however, only prion diseases are infectious. Clumps of prion proteins alone are thought to be the infectious agent in such diseases, implying that infectivity is a special property of these proteins. But, despite extensive studies, researchers have discovered little more than this about the basis for the transmissibility of prions. Now, by swapping portions of one yeast protein with those of another, Osherovich *et al.*¹ have found hints to a possible mechanism, as they discuss in *Public Library of Science Biology*.

Yeast prions are known to be transmitted between yeast strains, along with the cellular cytoplasm, during cell fusion, and from

mother to daughter yeast cells during cell division. Two events are necessary to ensure that these protein clumps continue to be transmitted. First, they must grow, by causing normal prion proteins to take on a warped shape that favours their aggregation. And second, they must divide, generating new prion particles that can be passed to a new cell. This division, or replication, is thought to involve small clumps breaking off from the main mass, with the help of a 'chaperone' molecule^{2,3}. Transmission then probably occurs by diffusion. Thus, transmission efficiency is related to the efficiency of prion replication.

Most proteins do not replicate in this way, so what enables a prion protein to do so? To find out, Osherovich *et al.*¹ looked at [PSI⁺] — the name given to the prion form of the yeast protein Sup35, which normally functions in the synthesis of other proteins. The

Materials science

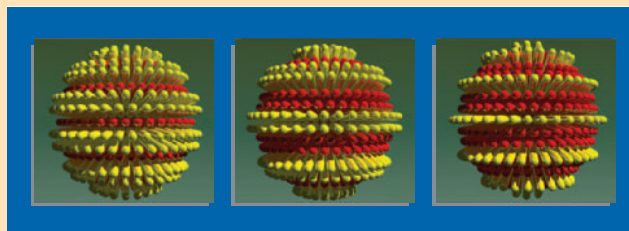
Variations on a golden core

Alicia M. Jackson *et al.* have looked at nanotechnology in the round, and describe what they believe to be a promising new class of structured material (*Nature Mater.* **3**, 330–336; 2004). They have examined nanoparticles that consist of a metal centre surrounded by a mixed-ligand shell. Using scanning tunnelling microscopy backed up by X-ray diffraction, they show that the shell is separated into ordered phases at the unprecedented scale of fractions of nanometres.

The particles were produced in a one-step synthesis involving the spontaneous assembly of ligands on nanosized gold cores. The ligand shell — a mixture of 1-octanethiol and mercaptopropionic acid —

separates into chemically distinct domains that are as small as 0.5 nm. These computer-generated representations show how the domains can form ripples that encircle the nanoparticles, and reveal the different ordering that stems from making the ligand shell with different molecular proportions of 1-octanethiol (yellow lobes) and mercaptopropionic acid (red). The nanoparticle diameter is 3.7 nm.

The principle works with cores of silver as well as gold. From this and other evidence, Jackson *et al.* argue that the phase separation that produces the different domains is not controlled by the characteristics of the core. Rather, after experimenting with ligand assembly on small gold



hemispheres of different diameters, they conclude that the ordering is primarily driven by surface curvature — how is not clear. But production of coated nanoparticles with tailored characteristics could involve varying the core curvature as well as the ligand type and mixture.

In looking into the properties of their nanoparticles, Jackson and colleagues tested the binding of

three proteins — cytochrome c, lysozyme and fibrinogen — to the ligand shell. The proteins didn't bind at all, and the authors consider that this failure to stick is probably down to the unique size and patterns of hydrophobic and hydrophilic areas on the shell. This 'nonspecific protein resistance' could be a major virtue in bioengineering applications.

Tim Lincoln

ability of Sup35 to form a prion depends on a region at one end of the protein, the amino-terminal end (Fig. 1). This well-studied 'prion domain' is rich in glutamine and asparagine amino acids—a property shared by all three confirmed yeast prion proteins. It also contains five-and-a-half repeats of a sequence of nine amino acids, which resemble five repeats that are seen in the only known mammalian prion protein, PrP. The glutamine/asparagine-rich region contributes to species specificity: the prion domains of Sup35 from different yeast species, which have different glutamine/asparagine-rich sequences, can aggregate in the same cell, but do so independently of one another⁴. The role of the repeats has been less clear, although they are involved in $[PSI^+]$ propagation.

Osherovich *et al.*¹ uncovered a role for the repeats while studying the prion-related properties of a region in New1, a putative prion protein that they had previously identified⁴ while searching the yeast genome for asparagine-rich sequences. In addition to an asparagine-rich region, New1 has one-and-a-half Sup35-like repeats. As the protein has no known function, it is difficult to show that it can behave as a prion (such a demonstration requires a protein's normal function to be disrupted by the prion form). But the authors did obtain some evidence for this when they created a new prion⁴—which they named $[NU^+]$ —by replacing the entire prion domain of Sup35 with that of New1.

Osherovich *et al.* now show that the asparagine-rich stretch alone of $[NU^+]$ causes aggregation—but that stable transmission of this prion also requires the repeats (Fig. 1). Deleting the repeats from Sup35 showed that they are also required for transmission of $[PSI^+]$. Surprisingly, however, propagation of $[PSI^+]$ seemed to require all five-and-a-half Sup35 repeats; four were not enough. That suggests that the New1 repeats might be better at supporting prion transmission (because just one-and-a-half such repeats do the trick).

The authors then showed that the repeats are interchangeable. They created a functional hybrid prion domain by fusing the asparagine-rich region of New1 to the Sup35 repeats (Fig. 1). This hybrid domain was 'species-specific' in that it interacted with $[NU^+]$ but not $[PSI^+]$ —confirming that the asparagine-rich region determines the specificity of aggregation, and showing that the repeats promiscuously confer transmissibility.

As a further demonstration of this promiscuity, Osherovich *et al.* designed another prion, based on the aggregation properties of tracts of glutamine amino acids. Tracts containing fewer than 35 consecutive glutamines do not aggregate readily, but those with longer stretches have a propensity to clump together in a manner that is proportional to the number of glutamines. This

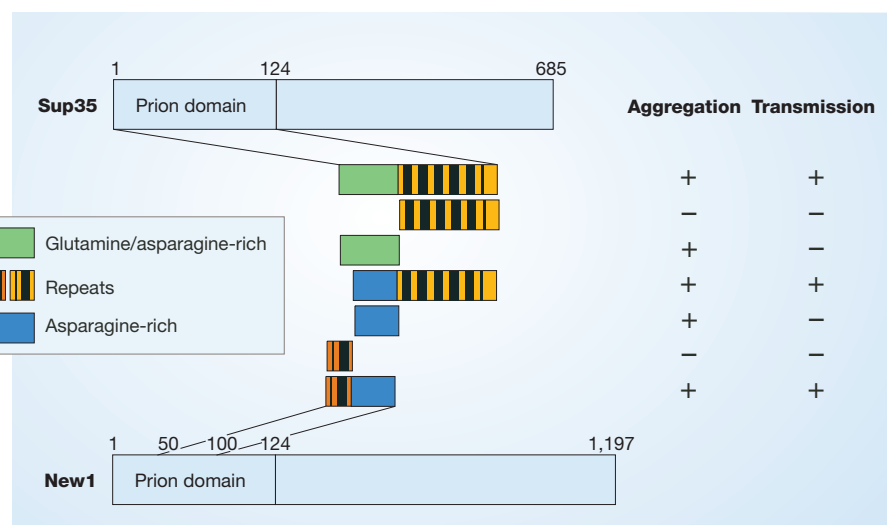


Figure 1 The modular properties of prion domains. The prion domains of the yeast Sup35 and New1 proteins contain several repeats of a particular amino-acid sequence (five-and-a-half repeats in Sup35, one-and-a-half in New1), along with either an asparagine-rich stretch or a tract that is rich in both asparagine and glutamine. Osherovich *et al.*¹ studied the effects of deleting one or other of these regions, and the results are shown here. Prion aggregation was measured by fusing the relevant domains to green fluorescent protein and looking for fluorescent foci. Transmission was tested by fusing the domains to the region of Sup35 that is involved in regulating protein synthesis, and assessing the stable propagation of prions among cells of a yeast population (as indicated by reduced termination of protein synthesis). The authors conclude that the asparagine- or glutamine/asparagine-rich region is needed for aggregation, and that both this region and the repeats are needed for transmission.

characteristic underlies the build-up of proteins seen in disorders such as Huntington's disease⁵. Osherovich *et al.* found that peptides containing 22 glutamines, with or without repeats, neither aggregated nor supported prion propagation. Peptides containing 62 glutamines readily aggregated—but required adjacent repeats from Sup35 to be transmissible. So the repeats did not cause aggregation but were necessary for transmission; aggregation was also a prerequisite for propagation.

In line with the view that the efficiency of transmission of yeast prions reflects prion replication, Osherovich *et al.* propose that the peptide repeats promote transmission by facilitating the chaperone-mediated division of aggregates. They suggest that the repeats either act as sites of interaction with chaperones or alter the conformation of aggregates in a way that increases their accessibility to chaperones. This scenario is simple, and provides a testable model for addressing questions on the properties of the repeats and the ability of prions that lack them to propagate stably. For example, do prions without repeats have other sequences that enable them to interact with chaperones? Or can they replicate on their own, as suggested by the discovery of a prion that apparently forms more fragile filaments, and so replicates without chaperones⁶? Yeast prions should prove useful for exploring such questions.

But can experiments with yeast provide insight into the infectivity of mammalian prions? The peptide repeats of mammalian

PrP are not essential for infectivity⁷. Intriguingly, however, a PrP repeat can functionally replace a Sup35 repeat to allow $[PSI^+]$ propagation⁸. And there are further parallels worth mentioning. PrP that contains more repeats than normal is associated with some familial prion diseases and with increased infectivity in mice⁶; extra repeats in Sup35 increase its propensity to become $[PSI^+]$ ⁹. Conversely, deleting repeats reduces both the infectivity of PrP and the appearance and transmissibility of $[PSI^+]$ ⁷⁻⁹. As Osherovich *et al.*¹ suggest, designing artificial yeast prions based on the aggregation properties of these proteins may provide a good model for studying aggregate-chaperone interactions in multicellular organisms.

Daniel C. Masison is in the Laboratory of Biochemistry and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0851, USA.

e-mail: masison@helix.nih.gov

- Osherovich, L. Z., Cox, B. S., Tuite, M. F. & Weissman, J. S. *PLoS Biol.* **2**, 0442–0451 (2004).
- Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Lieberman, S. W. *Science* **268**, 880–884 (1995).
- Paushkin, S. V., Kushnirov, V. V., Smirnov, V. N. & Ter-Avanesyan, M. D. *EMBO J.* **15**, 3127–3134 (1996).
- Santos, A., Chien, P., Osherovich, L. Z. & Weissman, J. S. *Cell* **100**, 277–288 (2000).
- Cummings, C. J. & Zoghbi, H. Y. *Annu. Rev. Genomics Hum. Genet.* **1**, 281–328 (2000).
- Crist, C. G., Nakayashiki, T., Kurahashi, H. & Nakamura, Y. *Genes Cells* **8**, 603–618 (2003).
- Weissmann, C. & Flechsig, E. *Br. Med. Bull.* **66**, 43–60 (2003).
- Parham, S. N., Resende, C. G. & Tuite, M. F. *EMBO J.* **20**, 2111–2119 (2001).
- Liu, J. J. & Lindquist, S. *Nature* **400**, 573–576 (1999).

Cancer

Homing in on tumour metastasis

Cancer Cell **5**, 365–374 (2004)

One of the deadliest hallmarks of tumour cells is their ability to spread to other organs. This process is often surprisingly specific; breast-cancer cells, for instance, frequently metastasize to the lungs and liver. Darren M. Brown and Erkki Ruoslahti have now identified a molecule in breast-tumour cells that allows them to home in on the lungs.

Using cultured mouse breast-cancer cells that can spread to the lungs and liver, Brown and Ruoslahti searched for proteins that bind specifically to lung blood vessels. This led to the discovery of the transmembrane protein metadherin. The authors found this protein on the surface of cells from several other cultured breast-cancer lines and from human breast tumours. But much less metadherin was present in normal breast tissue. They also showed that metadherin expression was sufficient to increase the spread of tumour cells to the lungs. Conversely, when metadherin expression was reduced or blocked with antibodies in the original cell line studied, fewer lung metastases formed when tested in a mouse model.

Although the molecule that metadherin recognizes in the lungs is unknown, this study suggests that drugs that target metadherin might be useful in preventing or treating lung metastases in breast-cancer patients.

Barbara Marte

Cell biology

Size control

J. Cell Sci. **117**, 1955–1960 (2004)

Some cell types grow to a certain length and then stop; others seem to lack length control. Margarita A. Kharitonova and Jury M. Vasiliev have found evidence that the ability to regulate length varies according to the structure of the cell's internal skeleton, rather than being rigidly specified by cell type.

A previous study showed that connective-tissue cells called fibroblasts maintain their normal length when compressed widthways in thin channels scratched on glass slides. By contrast, normally round epithelial cells take the opportunity to stretch out.

By using chemicals to disrupt the cells' cytoskeletons, Kharitonova and Vasiliev have now converted one cell shape into another. When fibroblasts lost their elongated microtubule and actin supports, they adopted a round shape and were then able to stretch out when placed in the channels. When epithelial cells shed their circular actin organization, they lengthened into a fibroblast-like shape but no longer extended further in the grooves.

The authors suggest that microtubules and actin fibres together control the cell's span. Both genes and external signals probably determine the shape that these parts of the cell's skeleton adopt.

Helen Pearson

Animal behaviour

On song in summer

Proc. R. Soc. Lond. B doi:10.1098/rspb.2004.2699 (2004)

Humpback whales confine their breeding and feeding to separate times and places — or so it was generally thought. Recordings of the mating songs of male whales show, however, that they continue to sing even when on their northern feeding grounds, indicating that mating is not restricted to their tropical overwintering grounds.



Christopher W. Clark and Phillip J. Clapham monitored the vocal displays of humpbacks (*Megaptera novaeangliae*) using sound recorders deployed on the sea floor at Georges Bank, a summer feeding area near Cape Cod, Massachusetts. Mating calls were detected for 25 consecutive days during May and June 2000.

Most breeding and singing happens during the winter, the authors say. But the possibility that mating also occurs in late spring could explain the occasional historical record of anomalously small fetuses in some whales caught at lower latitudes. As humpbacks gestate for almost a year, this could indicate that the females did not mate until after they had finished their northward trek.

Michael Hopkin

Chemistry

Built-in bypass

J. Am. Chem. Soc. **126**, 5154–5163 (2004)

Mark D. Erion *et al.* have designed a new type of liver-specific drug that might one day help to treat conditions such as hepatitis B and C and liver cancer. Called HepDirect prodrugs, the drug molecules

are based on nucleosides (DNA bases linked to sugar groups), which are known to halt viral replication and slow cell proliferation.

Nucleosides are activated by a series of enzymes that catalyse the sequential addition of phosphate groups. The addition of the first group is the most difficult, so HepDirect prodrugs are designed to bypass this by having the initial phosphate built in. When the drug comes into contact with the liver enzyme cytochrome P450, a protective chemical group is removed to expose the nucleoside–phosphate unit. Subsequent phosphate groups are then added by normal cell chemistry, and the drug becomes active.

Because cytochrome P450 is found mainly in the liver, the drugs' effects are restricted to this organ. Rats treated with the modified nucleoside display high levels of the activated nucleoside in their livers. Such improved targeting should enhance drug potency and also reduce the risk of side effects.

Helen R. Pilcher

Atmospheric science

Dark reactions

Geophys. Res. Lett. doi:10.1029/2004GL019412 (2004)

From ship-borne data collection, S. S. Brown and colleagues conclude that reactions occurring during the night are as important as daytime chemistry in regulating ozone levels in the lower atmosphere.

Ozone is formed during reactions involving nitrogen dioxide (NO_2), which can be removed from the atmosphere when it is converted to nitric acid and dissolves in raindrops or directly into the ocean. Conversion of NO_2 to nitric acid (HNO_3) will, in general, reduce the production of ozone. During the day, nitric-acid synthesis from NO_2 relies on a photochemical reaction, but at night the nitrate radical (NO_3) and nitrogen pentoxide (N_2O_5) take over as important intermediates in the conversion process.

Brown and colleagues have now more clearly quantified the role of these nocturnal compounds in removing NO_2 from the atmosphere. On average, about one-third of nitric-acid production happens at night, they say; acid production is lowest at dusk and dawn, when neither daytime nor nighttime processes dominate atmospheric chemistry.

Brown *et al.* also report that nitrates can be destroyed at night when they react with hydrocarbons emitted by plants, or with dimethyl sulphide rising from the ocean. The consequence, they conclude, is that the overall efficiency of night-time removal of NO_2 could be on a par with that during the day.

Mark Peplow

Moulting arthropod caught in the act

A Cambrian fossil confirms that early arthropods shed their coats just as they do today.

Until now, the existence of ecdysis (moulting) in early arthropods has been based solely on inference. Here we describe a 505-million-year-old specimen of the Cambrian soft-bodied arthropod *Marrella splendens* that has been visibly preserved in the middle of the act of moulting. This specimen confirms that early arthropods moulted during growth, just as they do today.

Ecdysis is a fundamental process that is thought to characterize the clade Ecdysozoa, which encompasses all moulting animals, including arthropods, tardigrades, onychophorans, nematodes, nematomorphs, kinorhynchans and priapulids¹. It may be that creatures in other groups moult², but evidence for this is anecdotal. Although ecdysis seems to have been a common feature of these phyla as far back as the Cambrian period³, evidence for moulting during the Cambrian is circumstantial.

For example, a few Cambrian trilobite specimens have been recorded as being preserved in an exuvial configuration⁴, indicating that they had just moulted. Mineralized hard parts of the trilobite exoskeleton found next to similar, less well mineralized parts have been interpreted as a new exoskeleton emerging from the old one, the exuvia. This interpretation is accepted because, even though trilobites have been extinct for 250 million years, their classification as arthropods is not questioned, and all arthropods living today moult during growth. Still, because the soft-bodied moult of the trilobite is not preserved, the interpretation remains only an inference.

Direct evidence of moulting is provided by a specimen of the arthropod *M. splendens* (Walcott 1912) from the Middle Cambrian Burgess Shale of British Columbia. This specimen is preserved halfway through the act of moulting, with its cephalic shield and lateral spines still flexible, squeezing out through an ecdysial opening at the front of the head shield of the old exoskeleton. The antennae are already freed; the distal ends of the lateral spines, and the rest of the body, have yet to emerge from the stiff exuvia (Fig. 1a, b). This remarkable specimen provides visible proof that Cambrian arthropods did indeed moult.

Why have Cambrian arthropods not been caught in the act of moulting before? The answer is twofold. First, a moulting event like this would only be recorded in taphonomic conditions that preserved soft tissues, and most fossil faunas did not have such an environment. Second, a non-mineralized arthropod would have taken a very short time to emerge completely from

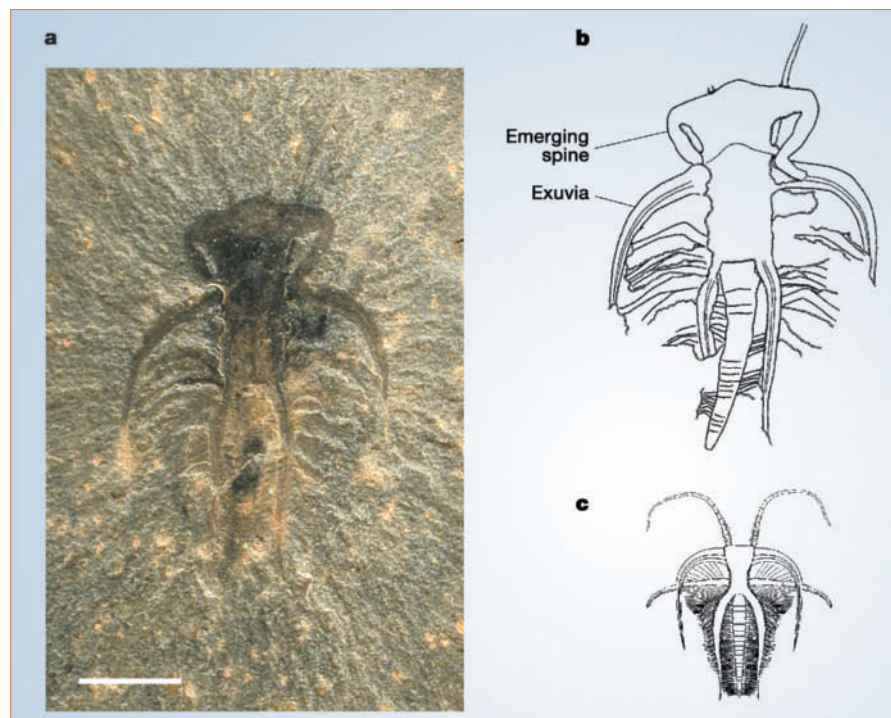


Figure 1 The oldest known fossil of an arthropod in the act of moulting: *Marrella splendens*, from the Middle Cambrian Burgess Shale of British Columbia, Canada. **a**, Specimen of *M. splendens* (ROM 56781) emerging and pulling out the flexible lateral spines from the old exoskeleton (exuvia). **b**, Camera lucida drawing of the same specimen. Scale bar for **a** and **b**, 5 mm. **c**, Reconstruction of *Marrella* (modified from ref. 8).

its exuvia; the duration would have been roughly comparable to the time that a similarly sized, non-mineralized lobster larva takes to moult (1–10 minutes; ref. 5) or to the 20 minutes that some cockroaches take to emerge from their exuviae⁶.

Why did we find evidence of this act in a Burgess Shale *M. splendens* rather than any other specimen? The Burgess Shale is justly famous for its exquisitely preserved fossils, which provide the best view of animals following the Cambrian evolutionary explosion of life roughly 520 million years ago. *M. splendens* is the most numerous arthropod in the Burgess Shale — more than 25,000 specimens have already been collected. If any Cambrian, soft-bodied arthropod is going to be preserved in the act of moulting, it is most likely to be *M. splendens* in the Burgess Shale.

Marrella is considered to be a basal arthropod because of its generalized morphology⁷. It has a head with two pairs of spines and two pairs of appendages; a trunk with up to 25 segments, each bearing a pair of biramous (two-branched) appendages; and a tiny last segment, or telson (Fig. 1c; ref. 8). The *Marrella* genus is included in its own small arthropod group, the Marrellomorpha, at the base of the cluster that includes crustaceans, trilobites and chelicerates⁹. So this instance of the early arthropod *Marrella*

splendens, preserved in the act of moulting 505 million years ago, confirms that ecdysis was occurring early in arthropod evolution.

Diego C. García-Bellido*†, Desmond H. Collins*

*Department of Palaeobiology, Royal Ontario Museum, Toronto, Ontario M5S 2C6, Canada
e-mail: desc@rom.on.ca

†Departamento de Paleontología, Facultad de Ciencias Geológicas, Universidad Complutense de Madrid, 28040 Madrid, Spain

1. Aguinaldo, A. M. A. et al. *Nature* **387**, 489–493 (1997).
2. Nielsen, C. *Zool. Scripta* **32**, 475–482 (2003).
3. Valentine, J. W. & Collins, A. G. *Evol. Dev.* **2**, 152–156 (2000).
4. Brandt, D. S. *Alcheringa* **26**, 399–421 (2002).
5. Matsuda, H., Takenouchi, T. & Yamakawa, T. *Fish. Sci.* **69**, 124–130 (2003).
6. Kunkel, J. G. *Biol. Bull.* **148**, 259–273 (1975).
7. Edgecombe, G. D. & Ramsköld, L. *J. Paleontol.* **73**, 263–287 (1999).
8. Whittington, H. B. *Geol. Surv. Can. Bull.* **209**, 1–24 (1971).
9. Wills, M. A. et al. in *Arthropod Fossils and Phylogeny* (ed. Edgecombe, G. D.) 33–105 (Columbia Univ. Press, New York, 1998).

Competing financial interests: declared none.

brief communications arising online
www.nature.com/bca

Optical media: Superluminal speed of information?

G. Nimtz (doi:10.1038/nature02586)

Reply: M. D. Stenner, D. J. Gauthier & M. A. Neifeld (doi:10.1038/nature02587)

Optical media

Superluminal speed of information?

Arising from: M. D. Stenner, D. J. Gauthier & M. A. Neifeld *Nature* **425**, 695–698 (2003)

The theory of special relativity limits signal velocity to the velocity of light in a vacuum. A faster-than-light (superluminal) signal velocity would violate causality^{1,2}. However, there are some experimental data and theoretical arguments that a causality violation does not necessarily happen if a signal velocity becomes superluminal³. Stenner *et al.*⁴ claim to have measured the speed of information in a fast-light optical medium by using a novel experimental set-up. The measured information (its front) travelled at a speed that did not exceed c (the velocity of light in vacuum). Their experimental result is correct but the interpretation is misleading because the information did not travel in the range of fast-light frequencies, as I explain here.

The authors transmitted a pulse with a carrier frequency of 389 THz through an optical medium having near this very frequency a narrow fast-light frequency band of 23 MHz. (The precise numbers of the relevant frequencies are not given.) The group refractive index n_g in this narrow band is said to be negative, with $n_g = -19.6$ resulting in a negative pulse speed. The information (generated by a sudden positive change of unity or a negative change of zero in the pulse amplitude, representing points of non-analyticity) was modulated at the pulse maximum. The modulation had a broad frequency band up to some gigahertz (that is, some nanoseconds in the time domain).

The complete frequency band of the signal is given by the carrier frequency plus the modulation frequency. It exceeds the extremely narrow fast-light frequency regime of $389 \text{ THz} \pm 11.5 \text{ MHz}$. Thus, the front of the information experienced a positive group refractive index $n_g = 3.38$ outside the small fast-light window according to its 3.3-nanosecond front delay. The front of the modulation, defined as information, travelled more slowly than light in a vacuum. The pulse without modulation has travelled with

superluminal speed through the fast-light frequency regime of the optical medium.

The new experiment does not provide an answer to the fundamental question of whether faster-than-light transport of information is possible.

Günter Nimtz

Universität zu Köln, II Physikalisches Institut,
50937 Köln, Germany
e-mail: g.nimtz@uni-koeln.de
doi:10.1038/nature02586

1. Sexl, R. & Urbantke, H. *Relativity, Groups, Particles* (Springer, Wien/New York, 2001).
2. Fayngold, M. *Special Relativity and Motions Faster than Light* (Wiley, Weinheim, 2002).
3. Nimtz, G. *Progr. Quant. Electr.* **27**, 417–450 (2003).
4. Stenner, M. D., Gauthier, D. J. & Neifeld, M. A. *Nature* **425**, 695–698 (2003).

Stenner et al. reply — It was most natural to present our findings¹ in the time domain, although there is an equivalent explanation in the frequency domain if Fourier analysis is used. Nimtz² uses a frequency-domain argument to assert that our data with gaussian pulses (Fig. 1b in ref. 1) are consistent with the group-velocity approximation, whereas the data we used to measure the information velocity (Fig. 2 in ref. 1) are not; however, we disagree with his analysis.

For the gaussian pulses, Fourier analysis reveals that the spectrum is a gaussian function centred on the optical carrier frequency and hence has small tails extending to infinitely high and low frequencies. Based on the time–frequency relation, a measurement of the pulse can be conducted in the time domain with a resolution that is much less than the full width at half-maximum (FWHM) of the pulse because of this infinite frequency extent. It is for this reason that we are able to measure the arrival time of information with a resolution better than about 1 nanosecond.

To determine whether a pulse has too broad a spectrum for the group-velocity

approximation to be valid, it is customary to compare the FWHM of the spectrum to the spectral region of anomalous dispersion. The appropriate measure for our experiment is the spacing between the gain peaks (23 MHz), which is much greater than the spectral width of about 1.1 MHz of the gaussian pulses, indicating that the approximation should be valid.

For the information-bearing pulses shown in Fig. 2a of ref. 1, the 10–90% fall-time, t_f , of the waveform corresponding to symbol ‘0’ is the fastest feature in our data, measured as 47 nanoseconds. From this value, the spectral width (FWHM) of our pulses is given by $0.35/t_f = 7.4 \text{ MHz}$. Therefore, both the gaussian and information-bearing pulses have FWHM spectral widths that are contained within the 23-MHz-wide region of anomalous dispersion, in contradiction to Nimtz’s claim.

Nimtz also contests that the group refractive index over a broad frequency range is positive and equal to 3.38, but we disagree. Although the group index is positive in the narrow spectral regions centred on each of the gain lines, the group index is negative over a large spectral region that lies beyond the gain lines.

We stress that most of the spectral power falls within the primary region of anomalous dispersion for all pulse shapes. However, it is true that the power in the tails of the information-bearing pulses is larger than that in the tails of the gaussian pulses. It is the subtle reshaping of the spectrum (both in the central part and tails) by the medium that gives rise to the subluminal information transfer shown in Fig. 2 of ref. 1.

Michael D. Stenner*, **Daniel J. Gauthier***,
Mark A. Neifeld†

*Duke University, Department of Physics, Box
90305, Durham, North Carolina 27708, USA
e-mail: gauthier@phy.duke.edu

†Department of Electrical and Computer
Engineering, The Optical Sciences Center,
University of Arizona, Tucson, Arizona 85721, USA
doi:10.1038/nature02587

1. Stenner, M. D., Gauthier, D. J. & Neifeld, M. A. *Nature* **425**, 695–698 (2003).
2. Nimtz, G. *Nature* doi:10.1038/nature02586 (2004)

Adult pancreatic β -cells are formed by self-duplication rather than stem-cell differentiation

Yuval Dor, Juliana Brown, Olga I. Martinez & Douglas A. Melton

Department of Molecular and Cellular Biology and Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

How tissues generate and maintain the correct number of cells is a fundamental problem in biology. In principle, tissue turnover can occur by the differentiation of stem cells, as is well documented for blood, skin and intestine, or by the duplication of existing differentiated cells. Recent work on adult stem cells has highlighted their potential contribution to organ maintenance and repair. However, the extent to which stem cells actually participate in these processes *in vivo* is not clear. Here we introduce a method for genetic lineage tracing to determine the contribution of stem cells to a tissue of interest. We focus on pancreatic β -cells, whose postnatal origins remain controversial. Our analysis shows that pre-existing β -cells, rather than pluripotent stem cells, are the major source of new β -cells during adult life and after pancreatectomy in mice. These results suggest that terminally differentiated β -cells retain a significant proliferative capacity *in vivo* and cast doubt on the idea that adult stem cells have a significant role in β -cell replenishment.

The literature on pancreatic β -cells and islets of Langerhans is replete with studies suggesting various mechanisms for β -cell homeostasis and regenerative repair. Early studies on patterns of [^3H]thymidine incorporation indicated that adult pancreatic endocrine cells belong to a class of tissues that could be maintained by the self-duplication of differentiated cells^{1–3}. More recent immunohistochemical observations suggest a stem-cell origin for islet cells, including insulin-expressing β -cells⁴. It has been proposed that these adult pancreatic stem or progenitor cells reside in the epithelium of pancreatic ducts^{5,6}, inside islets⁷ or in the bone marrow⁸. Others have suggested that β -cells form in the adult by transdifferentiation of pancreatic acinar cells⁹, islet cells that express hormones other than insulin¹⁰, or splenocytes¹¹. In addition to explaining the formation of new β -cells within existing islets, it has also been suggested that whole new islets form (islet neogenesis) by clustering of new β -cells that are derived from stem cells^{5,12,13}. However, all of these models and suggestions are, for the most part, based on the interpretation of static histological data rather than direct lineage analysis¹⁴.

We developed a simple method for distinguishing stem-cell-derived β -cells from the progeny of pre-existing β -cells. Fully differentiated β -cells, defined here as post-natal cells transcribing the insulin gene, are heritably labelled in transgenic mice with a tamoxifen-inducible Cre/lox system ('pulse'). The label is the expression of the human alkaline phosphatase protein, which can be detected by a histochemical stain. After a long period, during which turnover occurs ('chase'), β -cells are examined for the presence of the label. Cells generated after the pulse are labelled if and only if they are the progeny of pre-existing (labelled) β -cells; new β -cells derived from any non- β source, including stem cells, are not labelled. Different models of β -cell dynamics have distinct predictions regarding the frequency and distribution of labelled β -cells within pancreatic islets after a chase period (Fig. 1a). New islets derived entirely from stem or progenitor cells—that is, cells not yet transcribing the insulin gene—would not contain labelled β -cells. In addition, if stem cells replenish β -cells within existing islets, the frequency of labelled β -cells should gradually decrease. By contrast, if new β -cells are derived by self-duplication, the frequency of labelled β -cells within islets should remain constant.

We generated a transgenic mouse strain in which the insulin promoter¹⁵ drives the expression of tamoxifen-dependent Cre recombinase¹⁶ (RIP-CreER). In this strain, the Cre-oestrogen-receptor (ER) fusion gene is expressed only in pancreatic β -cells, but the CreER protein is excluded from the cell nucleus. Tamoxifen injection results in a rapid and transient (about 48 h (ref. 17)) nuclear translocation of the CreER protein, which permits Cre-mediated recombination. As a readout for Cre activity we used a reporter strain (Z/AP)¹⁸ in which Cre-mediated removal of a stop sequence results in the constitutive and heritable expression of the gene encoding human placental alkaline phosphatase (HPAP) (Fig. 1b). Thus, a pulse of tamoxifen to bigenic RIP-CreER;Z/AP mice leads to HPAP expression in insulin-expressing cells present at the time of injection, as well as their progeny.

Control experiments with RIP-CreER;Z/AP mice were per-

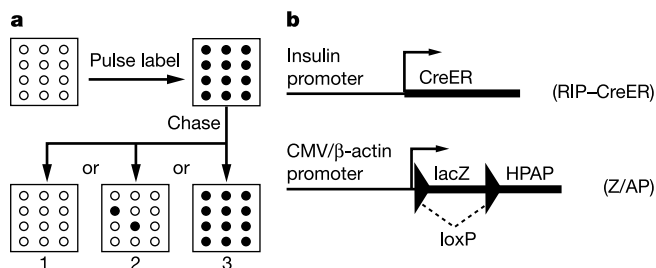


Figure 1 A pulse-chase system for determining the origin of β -cells. **a**, Predictions from different models of β -cell maintenance. Squares represent islets; open and filled circles represent unlabelled and labelled (HPAP⁺) β -cells, respectively. In this example, the initial labelling efficiency of β -cells is 100%. After a chase period, entirely stem-cell-derived new islets would contain no labelled cells (model 1). Within existing islets, maintenance by stem cells predicts a gradual decrease in the fraction of labelled β -cells (model 2), whereas maintenance by self-duplication predicts that the fraction of labelled β -cells remains constant (model 3). **b**, Transgenic mice. Tamoxifen injection of RIP-CreER;Z/AP bigenic mice results in a transient nuclear translocation of CreER, leading to removal of the loxP-flanked lacZ (serving as a stop sequence) and the permanent, heritable expression of the HPAP reporter gene. CMV, cytomegalovirus.

formed to assess the specificity of labelling after tamoxifen injection. HPAP⁺ cells are found only within islets (Fig. 2a); ducts, acini and blood vessels are not labelled (Fig. 2a). Double immunofluorescent analysis confirmed that ductal and acinar cells (detected by the DBA (*Dolichos biflorus* agglutinin) lectin and expressing amylase, respectively) do not express HPAP (Fig. 2c). Within islets, every HPAP⁺ cell ($n > 1,000$) expresses insulin (Fig. 2b). Non- β islet cells (those expressing glucagon, somatostatin or pancreatic polypeptide) were all HPAP⁻, as would be expected from an insulin-promoter-driven transgene (Fig. 2c). The frequency of HPAP⁺ cells depended on the tamoxifen dose (Fig. 2d). Thus, tamoxifen injection leads to a highly specific labelling of only differentiated β -cells (cells transcribing the insulin gene and positive for insulin by immunohistochemistry).

We injected tamoxifen into 24 RIP-CreER;Z/AP mice at 6–8 weeks of age, using a protocol (five injections of 4 mg tamoxi-

fen) that resulted in HPAP expression in about 30% of β -cells. Eight mice were killed 2 days after the last injection (the pulse group) and the remaining 16 mice were killed 2.5 ($n = 1$), 4 ($n = 2$), 6 ($n = 5$), 9 ($n = 4$) or 12 ($n = 4$) months later (the chase groups). Note that the 12-month time point represents about one-half of a mouse's lifespan. Sections of the pancreas were stained for insulin to identify β -cells and for HPAP to identify the cells that transcribed the insulin gene at time 0 (the pulse) and their daughters (see Fig. 1).

No new islets are formed during adult life

To identify whether entirely new islets were generated during the chase period, we compared the number of islets that contained labelled β -cells in the pulse and chase. Islets derived from stem cells (cells not expressing insulin at the time of pulse) would be completely HPAP⁻ in the chase pancreases. All islets analysed

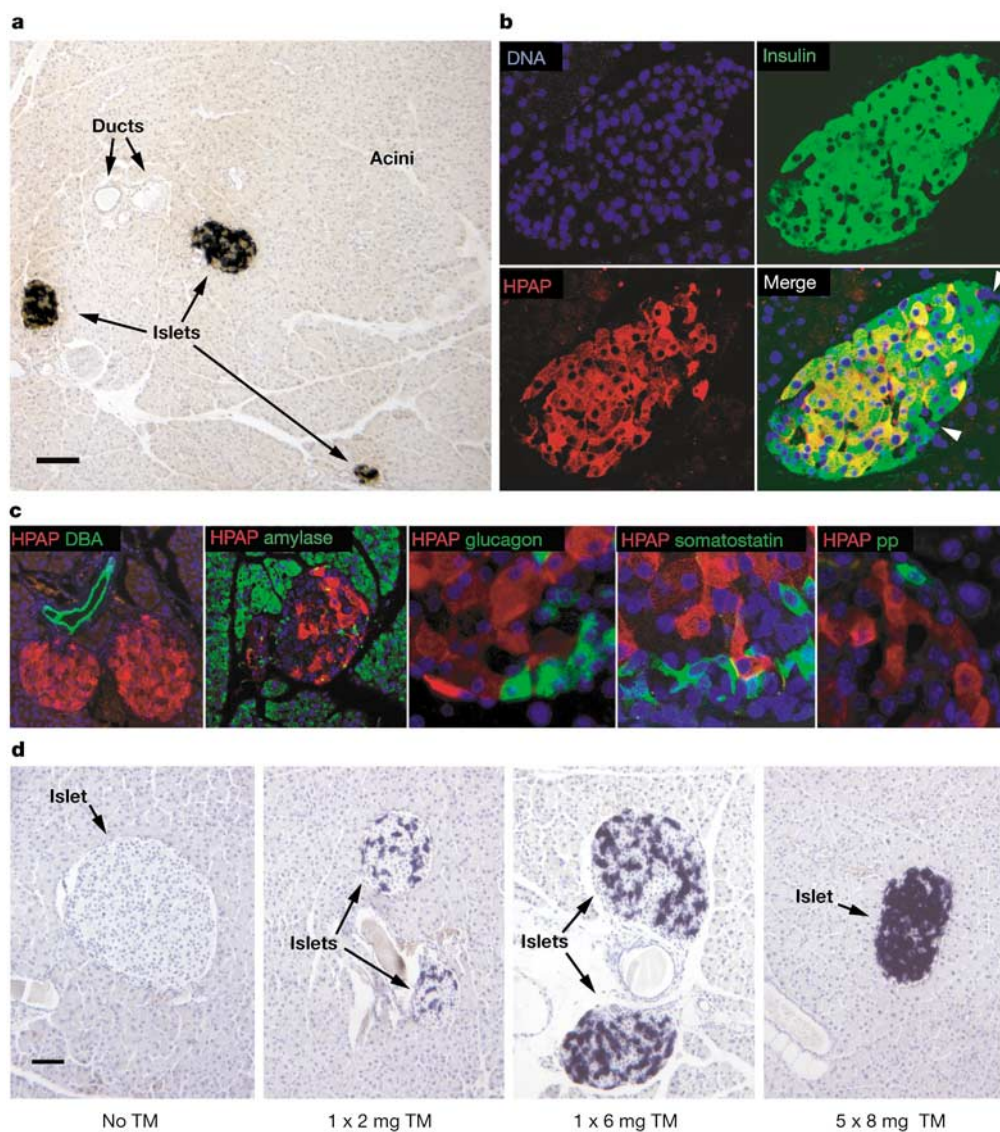


Figure 2 Specificity and dose dependence of recombination. **a**, HPAP expression is specific to islets. A section from the pancreas of a RIP-CreER;Z/AP mouse injected with tamoxifen 2 days before sacrifice and stained for insulin (brown) and HPAP (blue). HPAP⁺ cells are found in islets and are absent from acini and ducts. Original magnification, $\times 50$. Scale bar, 100 μ m. Note that co-staining for insulin and HPAP results in a dark blue to black colour. **b**, HPAP labelling within islets is specific to β -cells. A confocal image of a section similar to **a**, stained for DNA (4,6-diamidino-2-phenylindole (DAPI); blue), insulin (green) and HPAP (red). All HPAP⁺ cells also express insulin (orange in the lower-right

panel). Note that insulin-negative islet cells (for example, glucagon-positive cells) do not express HPAP (arrowheads). Original magnification, $\times 400$. **c**, HPAP⁺ cells (red) do not express non- β -cell markers (green). DNA is stained with DAPI (blue). Original magnification for the DBA and amylase staining, $\times 200$; for the glucagon, somatostatin and pp staining, $\times 400$. **d**, The degree of HPAP labelling in islets is dependent on the dose of tamoxifen (TM). Shown are sections of the pancreas from adult RIP-CreER;Z/AP mice, injected with the indicated dose of tamoxifen. Slides were stained for HPAP (dark blue) and counterstained with haematoxylin. Original magnification, $\times 100$. Scale bar, 50 μ m.

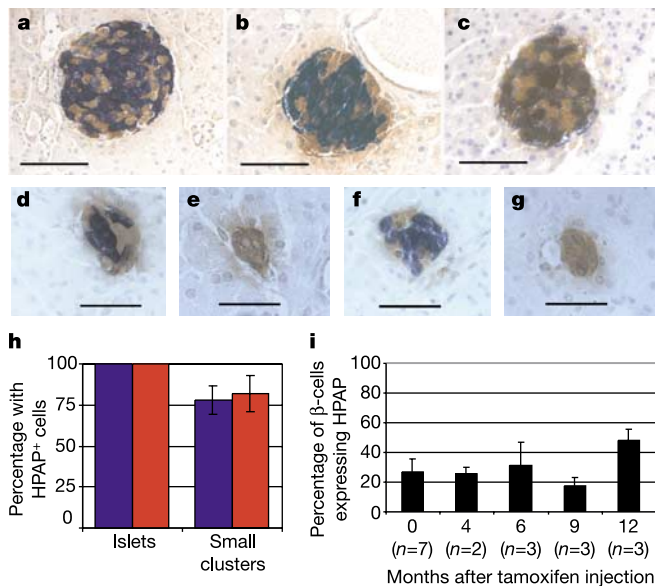


Figure 3 Analysis of islets and β -cells in pulse–chase experiments. **a–g**, Staining for HPAP (dark blue) and insulin (brown). **a–c**, Typical islets from mice killed immediately after tamoxifen injection (**a**), 4 months later (**b**) or 1 year later (**c**). Note that the HPAP stain obscures the brown insulin stain in double-positive cells. **d–g**, HPAP⁺ and HPAP[−] clusters of β -cells from mice immediately after the pulse (**d**, **e**) and in mice after a 6 month chase (**f**, **g**). Shown are clusters containing HPAP⁺ β -cells (**d**, **f**) and clusters containing only unlabelled β -cells (**e**, **g**). Original magnification, $\times 400$. Scale bars, 50 μ m (**a–c**), 25 μ m (**d–g**). **h**, Graph summarizing the frequency of HPAP⁺ islets and small clusters (containing fewer than 10 β -cells) in pulse–chase experiments with normal adult mice. Error bars represent standard deviations for all animals analysed ($n = 8$ for pulse, $n = 16$ for all chase experiments combined). Blue, pulse; red, chase. **i**, Graph summarizing the frequency of β -cells expressing HPAP in adult RIP–CreER;Z/AP mice after tamoxifen injection. Bars represent the average percentage of HPAP⁺ β -cells per mouse. Error bars represent standard deviations for all animals analysed for that time point; n is the number of mice.

($n = 485$ for the pulse group, $n = 744$ for the chase group; islet defined as a group of more than 10 β -cells) contained numerous HPAP⁺ β -cells. The finding that all islets in the chased mice contained β -cells that were present at the time of tamoxifen injection (or the progeny of such cells; Fig. 3a–c, h) suggests that no new islets are formed by stem or progenitor cells during adult life.

Given this unexpected result we extended our analysis to small,

scattered clusters of 1–10 β -cells. The cells in these structures are often interpreted as newly formed, stem-cell-derived β -cells caught in the process of coalescing into larger or mature islets^{9,19–23}. Such a model predicts that clusters found after a long chase should contain only HPAP[−] cells. In the pulse group, $77.8 \pm 8.4\%$ of the small clusters analysed ($n = 1,268$) contained HPAP⁺ β -cells, a frequency that was expected given the small size of the clusters and the efficiency of tamoxifen-dependent labelling. In the chase group, $81.8 \pm 10.9\%$ of the clusters analysed ($n = 1,423$) contained HPAP⁺ β -cells (Fig. 3d–h). Thus, there was no evidence that new clusters are generated from non- β -cells over the course of 12 months. We conclude, contrary to the commonly held view, that small clusters of β -cells do not represent stem-cell-derived islets. Rather, they might represent static mini-islets and/or disintegrating old islets. Taken together, these results point to the conclusion that no new islets are formed by stem or progenitor cells during adult life in the mouse (ruling out model 1 in Fig. 1a). The islets present in an old or middle-aged mouse are derived from islets that were present in ‘teenage life’ (2 months of age).

β -cells are formed by self-duplication of pre-existing β -cells

We next analysed these mice at the level of single cells, as opposed to islets, to examine β -cell maintenance. Sections of pulse and chase pancreases were stained for HPAP and insulin by double immunofluorescence, and scored for the percentage of β -cells that expressed HPAP⁺. The percentage of labelled β -cells, representing pre-existing β -cells or their progeny, remained stable in the chased mice (pulse, 4,759 β -cells counted from 7 mice, 27.4% HPAP⁺; chase, 8,988 β -cells counted from 11 mice, 29.54% HPAP⁺). Even 12 months after the pulse there was no indication of dilution of the labelled population (Fig. 3i). This result suggests that within established islets, β -cells are either post-mitotic (see below) or derive from the replication of pre-existing β -cells. The results are not consistent with β -cells forming from stem or progenitor cells (ruling out model 2 in Fig. 1a).

A trivial explanation to these results would be that very little β -cell turnover occurred during the chase. In such a model, new stem-cell-derived β -cells, even though HPAP[−], might escape detection because of their low rate of formation. On the contrary, fluorescence-activated cell sorting analysis of dissociated mouse pancreases shows that the absolute number of β -cells increases about 6.5-fold between 3 and 12 months of age ($(3.3 \pm 0.87) \times 10^5$ for 3-month-old mice, $n = 5$; $(2.2 \pm 0.9) \times 10^6$ for 10–12-month-old mice, $n = 4$) (Fig. 4). This means that at least 85% of β -cells present at 12 months of age were formed during the previous 9 months. Taking into account a conservative estimation of $t_{1/2} = 3$ months for adult rodent β -cells¹³, we estimate that 98% of β -cells

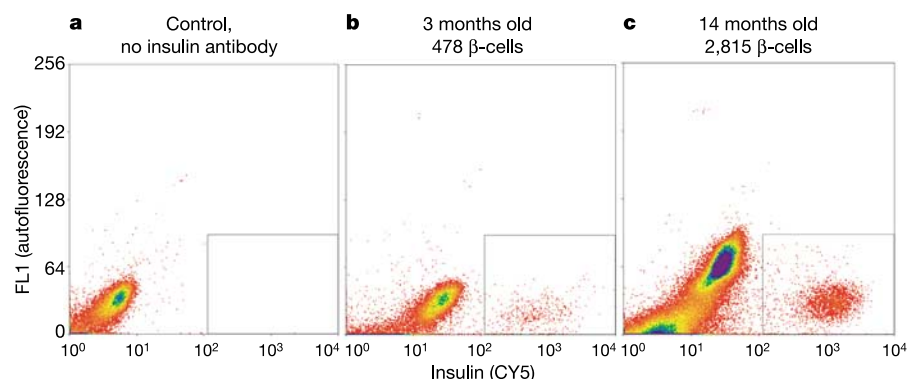


Figure 4 Total number of β -cells increases with age. Representative flow cytometric analysis of dissociated mouse pancreases stained for insulin expression. Shown are 0.1% of total pancreatic cells (**a**, **b**, 3-month-old nulliparous ICR female, 19,379 cells shown; **c**, 14-month-old nulliparous ICR female, 114,087 cells shown). **a**, No anti-insulin antibody added. The x axis is CY5 fluorescence (reflecting insulin staining); the y axis is autofluorescence. Insets, insulin-positive β -cells.

present in an animal chased for 9 months were born after the pulse. A 'static' interpretation of the pulse–chase results is therefore untenable and we conclude that new β -cells are generated in the adult mouse primarily by the replication of pre-existing β -cells.

β -cell regeneration

Although the above experiments do not provide any evidence for the existence of stem or progenitor cells in the normal growth and maintenance of β -cells, it is possible that stem or progenitor cells do participate in other circumstances. For instance, an emerging concept in the field of adult stem cells is that of facultative stem cells, namely cells that acquire a stem cell character only after stimulation^{24,25}. In the pancreas, such cells have been proposed to reside in ducts, acini or inside islets⁴. After injury, these cells are thought to first proliferate, then differentiate into β -cells and migrate to form new islets or replenish existing islets. To test this idea we used partial pancreatectomy (Px), an injury model in which regeneration has previously been documented^{14,5,12,26–29}. We first verified that Px did indeed result in the generation of new β -cells. Bromodeoxyuridine (BrdU) was administered for 1 week to mice that had undergone partial (70%) Px or had been sham operated. In agreement with previous reports^{5,26,27}, Px mice had a higher frequency of BrdU incorporation in all pancreatic cells (islets, acini, ducts and blood vessels) than sham-operated mice. This included a 2.6-fold increase in the frequency of BrdU⁺ β -cells after Px, indicative of β -cell regeneration (12.3% in Px versus 4.8% in sham-operated mice; $n = 7,388$ and 5,311 β -cells, respectively; Fig. 5a).

We injected tamoxifen into adult RIP–CreER;Z/AP mice ($n = 2$), and 4 days later 70% of their pancreas was removed. The mice were killed 2 months after the operation; the remnant of the pancreas, as well as the previously excised and fixed portion, were stained for insulin and HPAP. All islets examined in the remnant portion ($n = 126$) and in the excised portion ($n = 111$) contained numerous HPAP⁺ β -cells, showing that the islets found 2 months after Px existed before the operation. Furthermore, as observed during normal growth and adult life (Fig. 3), the frequency of clusters (groups of 1–10 β -cells) containing HPAP⁺ cells in the remnant portion (87.3%, $n = 276$) was similar to that in the excised portion (85.5%, $n = 325$), indicating that the clusters might not represent coalescing stem-cell-derived β -cells. These results suggest that, even after Px, no new islets are generated from stem cells. Furthermore, the frequency of HPAP⁺ β -cells within islets did not decrease after Px (data not shown), arguing against a significant contribution from non- β -cells or stem cells to the β -cells and islets found after pancreatectomy.

We wished to analyse directly the origin of new β -cells after Px, regardless of turnover rates or variations in labelling efficiency. To accomplish this we injected a high dose of tamoxifen (five injections of 8 mg) to another group of RIP–CreER;Z/AP mice ($n = 5$), and 14 days later performed 70% Px. BrdU was administered continuously for 2 weeks after the operation to label dividing cells, after which the mice were killed and their pancreases triple-stained for BrdU, insulin and HPAP (Fig. 5b). β -cells born after Px should express insulin and be BrdU⁺. If these new β -cells are derived from stem cells, they should be HPAP[−]; if they are the product of β -cell self-duplication, they should be HPAP⁺. The data show that new β -cells were HPAP⁺ in a frequency that precisely reflected the efficiency of recombination: 32 of 52 (61.5%) BrdU⁺ β -cells were HPAP⁺; 266 of 436 (61%) total β -cells counted were HPAP⁺. In other words, the results predict that if the labelling efficiency of β -cells were 100%, all new β -cells would have been HPAP⁺; that is, they would have come from β -cells existing at the time of Px. It is theoretically possible that a putative 'stem cell' directly differentiates into a β -cell without cell division—that is, without BrdU incorporation. However, the absence of HPAP dilution in the chase experiments described above is inconsistent with such a process contributing significantly to β -cell number. This result further

suggests that β -cells are derived from β -cells, not from stem cells.

Discussion

We conclude that pre-existing β -cells are the major source for new β -cells during adult life, as well as during regeneration from partial pancreatectomy (Fig. 1a, model 3). This conclusion challenges the view that adult stem cells (undifferentiated cells capable of self-renewal and differentiation) are significant in β -cell homeostasis. Instead, we submit that terminally differentiated β -cells retain a proliferative capacity that can account for turnover and expansion throughout a mouse's life.

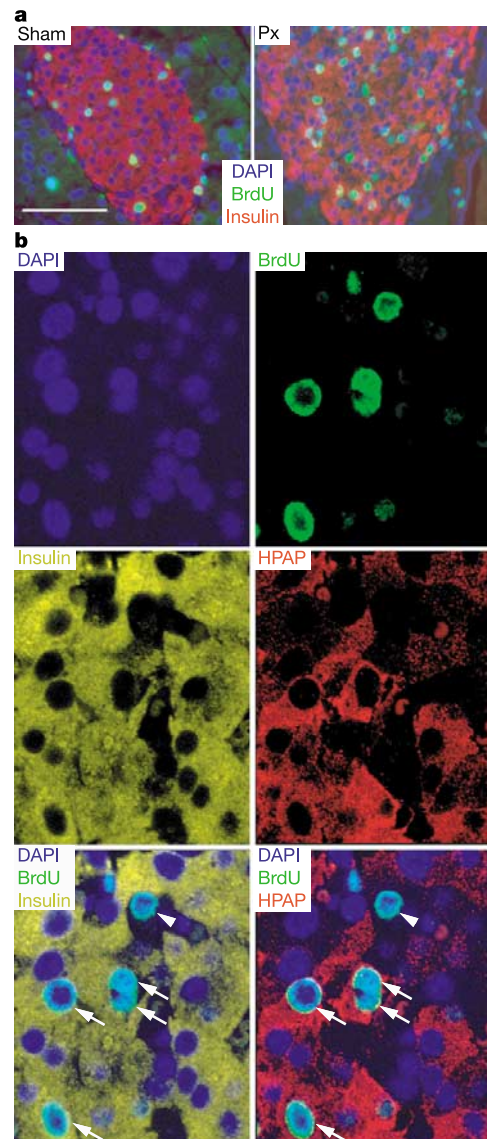


Figure 5 Analysis of β -cells after pancreatectomy. **a**, Increased BrdU incorporation after partial pancreatectomy (Px). Shown are representative islets from sham-operated or pancreatectomized mice, 7 days after operation. BrdU was administered continuously between operation and killing. Staining for BrdU (green), insulin (red) and DNA (blue). Original magnification, $\times 400$. Scale bar, 50 μ m. **b**, Origins of new β -cells. Partial pancreatectomy was performed on tamoxifen-injected RIP–CreER;Z/AP mice, followed by BrdU administration for 2 weeks. Shown are confocal images of an islet stained for insulin (yellow), HPAP (red), BrdU (green) and DNA (blue). Four BrdU⁺ β -cells are also HPAP⁺ (arrows), indicating that they were generated by replication of existing β -cells. One BrdU⁺ cell is insulin-negative and HPAP[−] (arrowhead); this is probably an α or δ cell. Original magnification, $\times 650$.

There are several ways in which our data could be reconciled with previous reports claiming the existence of pancreatic stem cells. First, our experiments cannot prove the absence of stem cells, given the degree of variation in labelling efficiency inherent to the CreER system. It is possible that rare stem or progenitor cells exist but give rise to a small, physiologically unimportant fraction of β -cells¹⁷. A convincing demonstration of such putative stem cells could be provided by genetic lineage tracing, for example by using a duct cell promoter driving Cre recombinase. Second, it is formally possible that the insulin gene is transcriptionally active in stem cells. If so, these putative stem cells must be unipotent because HPAP⁺ cells gave rise only to β -cells. Even after a long chase, no other endocrine, exocrine or ductal cells were HPAP⁺. In effect, such putative unipotent stem cells would be indistinguishable from β -cells. Third, facultative pancreatic stem cells might exist and be activated in response to specific insults other than pancreatectomy^{24,30}. Fourth, the vast majority of Px and regeneration studies were done with rats rather than mice and it is possible that these species differ in the mode of regeneration. Last, β -cells could transiently dedifferentiate (and possibly acquire stem cell markers) before generating new β -cells¹². A similar process is thought to occur during angiogenesis, in which differentiated endothelial cells give rise to 'activated endothelium', which then proliferates and organizes into new blood vessels³¹. However, there is currently no evidence for such a mechanism in the pancreas.

It has been suggested that new lobes, including new islets, are constantly generated in the adult pancreas^{13,32}. In contrast, the data presented here support the idea that the number of islets during adult life is fixed, with homeostatic responses (for example, β -cell expansion) being performed by differentiated cells residing in these structural units. Indeed, it was recently reported that average islet size, but not islet density, increases with age^{33–35}. Our studies indicate that there is a time during embryonic development, or early postnatal life, at which the number of islets is set. The data also indicate that β -cell formation from undifferentiated precursors ceases at this time in that new β -cells that form subsequently come from pre-existing β -cells. Further experiments are required to determine when this switch occurs.

Our results emphasize the importance of using direct lineage tracing, rather than histological snapshots, to determine a cell of origin or the provenance of a given tissue. The method presented here can be adapted to assess the contribution of stem cells in other organs during growth, normal turnover and regeneration. In addition, it could be used to determine the cell of origin in cancer models.

One implication of this study concerns possible cell-based therapies for type I diabetes, in which β -cells are destroyed by an autoimmune attack. It points to the significant proliferative potential of differentiated β -cells *in vivo*, a capacity that might be exploited for expansion to a clinically useful mass. In addition, the results highlight the fact that embryonic stem cells are currently the only type of stem cell that is unquestionably capable of differentiation into β -cells. □

Methods

Mice

Z/AP reporter mice were a gift from C. Lobe¹⁸. The RIP–CreER construct was generated by fusing a 0.66-kilobase *SmaI*–*HindIII* fragment of the RIP2 promoter¹⁵ to a minimal hsp68 promoter (a gift from M. Gannon³⁶), and placing the chimaeric promoter upstream of CreER (a gift from A. McMahon¹⁶). Transgenic mice were created by injection of the construct into the pronucleus. Three founders were identified that, when crossed with Z/AP, showed β -cell-specific expression of HPAP (defined by immunohistochemical co-localization of HPAP and insulin). For this study we used two lines that showed efficient recombination. Both lines showed tamoxifen-dependent expression of HPAP. By 10 months of age, double transgenic mice that did not receive tamoxifen had occasionally one to five HPAP⁺ cells per islet, reflecting the cumulative lifetime leakage of the system. RIP–CreER;Z/AP double transgenic mice were generated by crossing single heterozygous transgenics. Tamoxifen (Sigma) was dissolved in corn oil at 20 mg ml^{−1} and injected intraperitoneally or subcutaneously twice a week. RIP–CreER;Z/AP mice treated with

tamoxifen remained euglycaemic and had normal pancreatic histology, indicating that tamoxifen injection, Cre-mediated recombination and HPAP expression had no adverse effects. For continuous labelling of dividing cells, BrdU (Sigma) was given in the drinking water at 0.8 mg ml^{−1}, and the water was changed every other day.

Analysis

Formalin-fixed, paraffin-embedded 6- μ m sections of the pancreas were used. We identified β -cells by immunostaining for insulin (guinea-pig anti-insulin (Dako), dilution 1:500) or for pdx1 (guinea-pig anti-pdx1 (a gift from C. Wright), dilution 1:500). Other pancreatic cell types were identified with the following antibodies: guinea-pig anti-glucagon (Linc; dilution 1:200), sheep anti-somatostatin (ARP; dilution 1:200), guinea-pig anti-pancreatic polypeptide (Linc; dilution 1:200) and rabbit anti-amylase (Sigma; dilution 1:200). Ducts were labelled with biotinylated DBA lectin (vector; dilution 1:200). HPAP⁺ cells were identified by incubating slides with alkaline phosphatase substrate as described^{17,18} or by immunostaining (rabbit anti-HPAP (Zymed), dilution 1:100, or goat anti-HPAP (Santa Cruz), dilution 1:100). BrdU⁺ cells were stained with biotinylated anti-BrdU antibodies (BrdU in-situ detection kit (Becton Dickinson), dilution 1:100). For both HPAP and BrdU immunostaining we performed antigen retrieval (Retrievagen A; Becton Dickinson) before incubation with primary antibody. Secondary antibodies conjugated with horseradish peroxidase, fluorescein isothiocyanate, rhodamine (Jackson) or CY5 (Molecular Probes) were used at 1:200 dilution. Flow cytometry was carried out with MoFlo (Cytomation). Pancreases were dissociated to a single-cell suspension essentially as described^{37,38} and cells were stained with guinea-pig anti-insulin antibody followed by a CY5-conjugated anti-guinea-pig antibody.

Received 30 December 2003; accepted 29 March 2004; doi:10.1038/nature02520.

- Messier, B. & Leblond, C. P. Cell proliferation and migration as revealed by radioautography after injection of thymidine-H³ into male rats and mice. *Am. J. Anat.* **106**, 247–285 (1960).
- Leblond, C. P. in *International Symposium on the Control of Cell Division and the Induction of Cancer* (ed. Congdon, C. C.) 119–150 (Nat'l Cancer Inst., Bethesda, Maryland, 1964).
- Tsubouchi, S., Kano, E. & Suzuki, H. Demonstration of expanding cell populations in mouse pancreatic acini and islets. *Anat. Rec.* **218**, 111–115 (1987).
- Bonner-Weir, S. & Sharma, A. Pancreatic stem cells. *J. Pathol.* **197**, 519–526 (2002).
- Bonner-Weir, S., Baxter, L. A., Schuppert, G. T. & Smith, F. E. A second pathway for regeneration of adult exocrine and endocrine pancreas. A possible recapitulation of embryonic development. *Diabetes* **42**, 1715–1720 (1993).
- Zajack, G., Arber, N., Schwartz-Arad, D. & Ariel, I. Streaming pancreas: islet cell kinetics. *Diabetes Res.* **13**, 121–125 (1990).
- Zulewski, H. *et al.* Multipotential nestin-positive stem cells isolated from adult pancreatic islets differentiate *ex vivo* into pancreatic endocrine, exocrine, and hepatic phenotypes. *Diabetes* **50**, 521–533 (2001).
- Ianus, A., Holz, G. G., Theise, N. D. & Hussain, M. A. *In vivo* derivation of glucose-competent pancreatic endocrine cells from bone marrow without evidence of cell fusion. *J. Clin. Invest.* **111**, 843–850 (2003).
- Lipsett, M. & Finegood, D. T. Beta-cell neogenesis during prolonged hyperglycemia in rats. *Diabetes* **51**, 1834–1841 (2002).
- Guz, Y., Nasir, I. & Teitelman, G. Regeneration of pancreatic beta cells from intra-islet precursor cells in an experimental model of diabetes. *Endocrinology* **142**, 4956–4968 (2001).
- Kodama, S., Kuhlreiber, W., Fujimura, S., Dale, E. A. & Faustman, D. L. Islet regeneration during the reversal of autoimmune diabetes in NOD mice. *Science* **302**, 1223–1227 (2003).
- Bonner-Weir, S. Life and death of the pancreatic beta cells. *Trends Endocrinol. Metab.* **11**, 375–378 (2000).
- Finegood, D. T., Scaglia, L. & Bonner-Weir, S. Dynamics of beta-cell mass in the growing rat pancreas. Estimation with a simple mathematical model. *Diabetes* **44**, 249–256 (1995).
- Slack, J. M. Developmental biology of the pancreas. *Development* **121**, 1569–1580 (1995).
- Hanahan, D. Heritable formation of pancreatic beta-cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes. *Nature* **315**, 115–122 (1985).
- Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K. & McMahon, A. P. Modification of gene activity in mouse embryos *in utero* by a tamoxifen-inducible form of Cre recombinase. *Curr. Biol.* **8**, 1323–1326 (1998).
- Gu, G., Dubauskaite, J. & Melton, D. A. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447–2457 (2002).
- Lobe, C. G. *et al.* Z/AP, a double reporter for cre-mediated recombination. *Dev. Biol.* **208**, 281–292 (1999).
- Wang, R. N., Kloppel, G. & Bouwens, L. Duct- to islet-cell differentiation and islet growth in the pancreas of duct-ligated adult rats. *Diabetologia* **38**, 1405–1411 (1995).
- Yamamoto, K. *et al.* Recombinant human betacellulin promotes the neogenesis of beta-cells and ameliorates glucose intolerance in mice with diabetes induced by selective alloxan perfusion. *Diabetes* **49**, 2021–2027 (2000).
- Waguri, M. *et al.* Demonstration of two different processes of beta-cell regeneration in a new diabetic mouse model induced by selective perfusion of alloxan. *Diabetes* **46**, 1281–1290 (1997).
- Paris, M., Bernard-Kargar, C., Berthault, M. F., Bouwens, L. & Ktorza, A. Specific and combined effects of insulin and glucose on functional pancreatic beta-cell mass *in vivo* in adult rats. *Endocrinology* **144**, 2717–2727 (2003).
- Rooman, I., Lardon, J. & Bouwens, L. Gastrin stimulates beta-cell neogenesis and increases islet mass from transdifferentiated but not from normal exocrine pancreas tissue. *Diabetes* **51**, 686–690 (2002).
- Alison, M. R. Liver regeneration with reference to stem cells. *Semin. Cell Dev. Biol.* **13**, 385–387 (2002).
- Oh, S. H., Hatch, H. M. & Petersen, B. E. Hepatic oval 'stem' cell in liver regeneration. *Semin. Cell Dev. Biol.* **13**, 405–409 (2002).
- Liu, Y. Q., Montanya, E. & Leahy, J. L. Increased islet DNA synthesis and glucose-derived lipid and amino acid production in association with beta-cell hyperproliferation in normoglycaemic 60% pancreatectomy rats. *Diabetologia* **44**, 1026–1033 (2001).
- Fong, I., Hulsink, I., Fortuna, R. & Silink, M. Mitotic activity and DNA synthesis of rat islet cells

- following partial pancreatectomy. *Pancreas* **7**, 453–459 (1992).
28. Finegood, D. T., Weir, G. C. & Bonner-Weir, S. Prior streptozotocin treatment does not inhibit pancreas regeneration after 90% pancreatectomy in rats. *Am. J. Physiol.* **276**, E822–E827 (1999).
29. Hardikar, A. A., Karandikar, M. S. & Bhonde, R. R. Effect of partial pancreatectomy on diabetic status in BALB/c mice. *J. Endocrinol.* **162**, 189–195 (1999).
30. Fausto, N. & Campbell, J. S. The role of hepatocytes and oval cells in liver regeneration and repopulation. *Mech. Dev.* **120**, 117–130 (2003).
31. Carmeliet, P. Angiogenesis in health and disease. *Nature Med.* **9**, 653–660 (2003).
32. Bunnag, S. C. Postnatal neogenesis of islets of Langerhans in the mouse. *Diabetes* **15**, 480–491 (1966).
33. Kushner, J. A. *et al.* Pdx1 restores beta cell function in Irs2 knockout mice. *J. Clin. Invest.* **109**, 1193–1201 (2002).
34. Skau, M., Pakkenberg, B., Buschard, K. & Bock, T. Linear correlation between the total islet mass and the volume-weighted mean islet volume. *Diabetes* **50**, 1763–1770 (2001).
35. Bock, T., Pakkenberg, B. & Buschard, K. Increased islet volume but unchanged islet number in *ob/ob* mice. *Diabetes* **52**, 1716–1722 (2003).
36. Wu, K. L. *et al.* Hepatocyte nuclear factor 3 β is involved in pancreatic beta-cell-specific transcription of the *pdx-1* gene. *Mol. Cell. Biol.* **17**, 6002–6013 (1997).
37. Freshney, R. I. *Culture of Animal Cells: A Manual of Basic Technique* (Wiley-Liss, New York, 2000).
38. Arkle, S., Lee, C. M., Cullen, M. J. & Argent, B. E. Isolation of ducts from the pancreas of copper-deficient rats. *Q. J. Exp. Physiol.* **71**, 249–265 (1986).

Acknowledgements We thank J. Dubauskaite for zygote injection, G. Kenty for help with fluorescence-activated cell sorting analysis, H. Hsieh for technical assistance, and S. Bonner-Weir, A. Regev and members of the Melton laboratory for discussions. D.A.M. is a Howard Hughes Medical Institute Investigator. Y.D. was supported by EMBO and JDRF postdoctoral fellowships.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.M. (dmelton@mcb.harvard.edu).

The central dusty torus in the active nucleus of NGC 1068

W. Jaffe¹, K. Meisenheimer², H. J. A. Röttgering¹, Ch. Leinert², A. Richichi³, O. Chesneau², D. Fraix-Burnet⁴, A. Glazenberg-Kluttig⁵, G.-L. Granato⁶, U. Graser², B. Heijligers¹, R. Köhler², F. Malbet⁴, G. K. Miley¹, F. Paresce³, J.-W. Pei⁷, G. Perrin⁸, F. Przygodda², M. Schoeller⁹, H. Sol⁸, L. B. F. M. Waters^{10,11}, G. Weigelt¹², J. Woillez⁸ & P. T. de Zeeuw¹

¹Leiden Observatory, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands

²Max Planck Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

³European Southern Observatory, Karl Schwarzschild Strasse 2, D-85748 Garching bei München, Germany

⁴Observatoire de Grenoble, BP 53, F-38041 Grenoble, Cedex 9, France

⁵Stichting ASTRON, 7990 AA Dwingeloo, Netherlands

⁶Astronomical Observatory of Padua, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy

⁷Kapteyn Astronomical Institute, Postbus 800, 9700 AV Groningen, The Netherlands

⁸Observatoire de Paris-Meudon, 92195 Meudon, France

⁹European Southern Observatory, Casilla 19001, Santiago, Chile

¹⁰Astronomical Institute "Anton Pannekoek", Univ. of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

¹¹Instituut voor Sterrenkunde, Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3001 Heverlee, Belgium

¹²Max-Planck-Institut für Radioastronomie, Auf dem Hugel 69, D-53121 Bonn, Germany

Active galactic nuclei (AGNs) display many energetic phenomena—broad emission lines, X-rays, relativistic jets, radio lobes—originating from matter falling onto a supermassive black hole. It is widely accepted that orientation effects play a major role in explaining the observational appearance of AGNs. Seen from certain directions, circum-nuclear dust clouds would block our view of the central powerhouse^{1,2}. Indirect evidence suggests that the dust clouds form a parsec-sized torus-shaped distribution. This explanation, however, remains unproved, as even the largest telescopes have not been able to resolve the dust structures. Here we report interferometric mid-infrared observations that spatially resolve these structures in the galaxy NGC 1068. The observations reveal warm (320 K) dust in a structure 2.1 parsec thick and 3.4 parsec in diameter, surrounding a smaller hot structure. As such a configuration of dust clouds would collapse in a time much shorter than the active phase of the AGN³, this observation requires a continual input of kinetic energy to the cloud system from a source coexistent with the AGN.

NGC 1068^{4,5} (Fig. 1) is a well studied Seyfert 2 type AGN at a distance of only 14.4 Mpc (megaparsec; 1 pc = 3.26 light yr). In contrast to quasars or Seyfert 1 galaxies, we view Seyfert 2 galaxies at an aspect angle along which the central energetic phenomena and particularly the hot central accretion disk seem to be blocked by dust¹. The largest single-dish telescopes provide a resolution of 100 milli-arcsec (mas) and fail to resolve the circum-nuclear dust structure (at 14.4 Mpc, 100 mas corresponds to 7 pc). The dust is heated by the hot accretion disk to temperatures between ~100 K and the dust sublimation temperature, 1,500 K, and thus radiates thermally between wavelengths $\lambda \approx 2$ and 30 μm , although dust absorption will strongly suppress the emergence of any radiation of wavelength $\lambda \leq 4 \mu\text{m}$ from Seyfert 2 galaxies.

To resolve the dust structures and thus measure their properties directly at $\lambda > 8 \mu\text{m}$, we used the mid-infrared interferometric instrument (MIDI)⁶ coupled to the European Southern Observatory's (ESO's) Very Large Telescope interferometer (VLT)⁷. The spatial resolution of the largest two-telescope configuration of the

VLT is ~ 10 mas at $\lambda = 10 \mu\text{m}$, and allows us to resolve thermal emission from dust to < 1 pc in NGC 1068. MIDI/VLTI operates as a classical astronomical Michelson interferometer. The light beams from each two of the VLT's unit telescopes are relayed by a series of

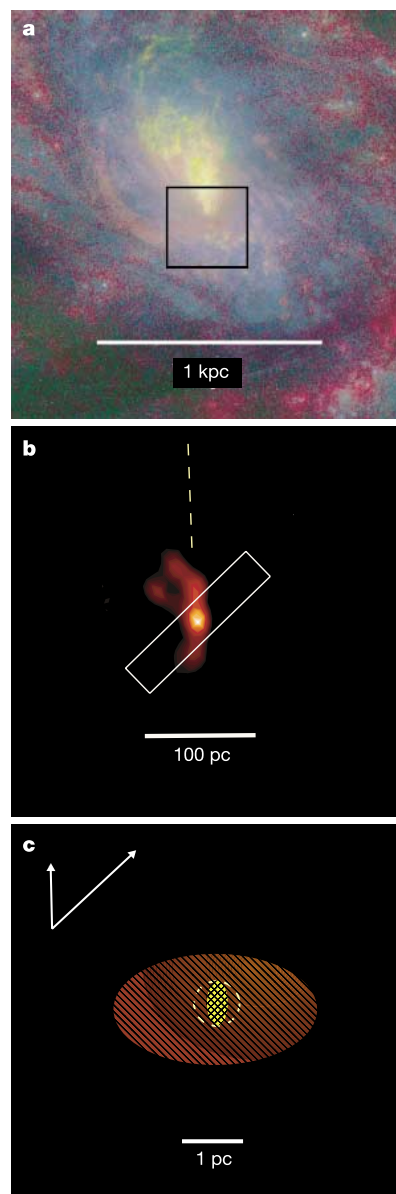


Figure 1 Images and model of emission from NGC 1068 at increasing magnification. **a**, Optical image of the central region of NGC 1068²⁰. This colour composite taken with the Hubble Space Telescope shows stellar light in blue, oxygen ionized by the active nucleus in yellow and ionized hydrogen in red. The black square centred on the dust-obscured nucleus marks the region of **b**. In **a–c**, north is at the top, and east to the left. **b**, Single-telescope acquisition image (deconvolved) of NGC 1068 taken by MIDI with a 8.7 μm filter penetrating the dust and showing the structures on arcsec scales. These are similar to those of ref. 9. Also shown is the position of the spectroscopic slit used in the interferometric observations. The spectra displayed in Fig. 2 are integrated over the central emission peak only. The dashed line indicates the source axis. **c**, Model dust structure in the nucleus of NGC 1068. It contains a central hot component (dust temperature $T > 800$ K, yellow) marginally resolved along the source axis. Its finite width and the dashed circle indicate that its east–west extent is currently undetermined. The much larger warm component ($T = 320$ K, red) is well resolved. Single hatching represents the averaged optical depth in the silicate absorption, $\langle \tau_{\text{SiO}} \rangle = 0.3$, while the higher value $\tau_{\text{SiO}} = 2.1$ is found towards the hot component (cross-hatched). The arrows indicate the projected orientation of the two baselines and the spatial resolution $\theta = \langle \lambda \rangle / 2B$, where B is the projected baseline.

mirrors, combined by a semi-reflecting mirror, dispersed by a NaCl prism to a spectral resolution of $\lambda/\delta\lambda \approx 30$, and finally registered on the infrared detector. MIDI's operation in the astronomical N band (8–13.5 μm) is perfectly suited to estimate the dust composition from the absorption lines of common dust minerals.

Figure 1b shows our single-telescope acquisition image of the nuclear region at 8.7 μm wavelength. The corresponding N-band slit spectrum of the source, as integrated over the beam width of 0.4 arcsec, is shown in Fig. 2a. Figures 2b and c show the spectra as seen by the two-telescope interferometer operating at two different projected baselines. We attribute the dip near 10 μm , visible on all three panels, to absorption by dust minerals.

As our longer baseline is essentially aligned with the north–south-oriented source axis as defined by the inner radio jet⁸, we

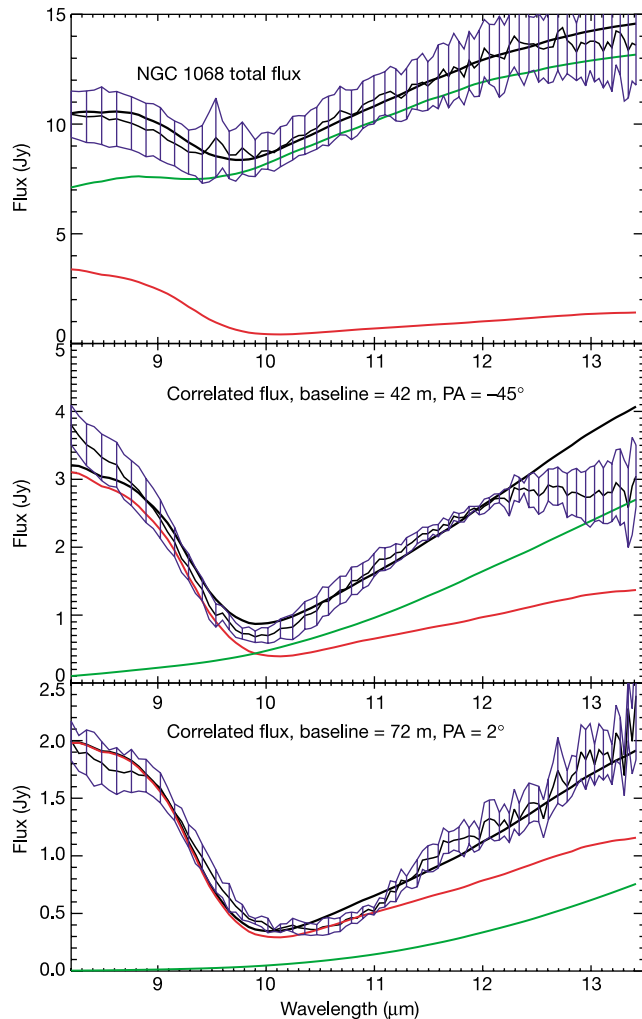


Figure 2 Observed MIDI spectra from the nucleus of NGC 1068 compared with two-component gaussian model predictions. Spectra were obtained during runs on 15 and 16 June, and 9 and 10 November, 2003. In each plot, the thin jagged line shows the observed flux as a function of wavelength, the blue hatched area shows its r.m.s. scatter in independent observations, and the thicker smooth line shows the model predictions. The red and green lines show the contribution of the hot and warm components, respectively. **a**, The single-telescope, non-interferometric MIDI flux-density spectrum. The absorption dip near 10 μm is caused primarily by ‘astronomically common’, olivine-type, silicate dust. **b**, The two-telescope interferometric spectrum from November 2003 with a projected baseline of $B = 42$ m oriented at an angle of 45° (clockwise from north), giving an effective resolution of ~ 26 mas. **c**, The interferometric spectrum from June 2003 with a projected baseline of 78 m, orientated 2° anticlockwise from north, giving a resolution of ~ 13 mas. Here the dip near 10 μm is probably caused by dust of aluminosilicate or other, non-olivine, composition.

decided to model the data with a minimum set of emitting components that are symmetric with respect to this axis—a procedure borrowed from the early days of VLBI radio astronomy. For each component, we assume blackbody radiation at a specific temperature and an elliptical gaussian brightness distribution. In order to reproduce the dust absorption feature already evident in our present (and previous^{9,10}) single-telescope observations, we additionally introduce silicate absorption of various strengths for each component. We assume an absorption profile derived from Infrared Satellite Observatory (ISO) observations of the centre of our Galaxy¹¹, which is consistent with, but of higher signal/noise than, existing ISO measurements of NGC 1068¹⁰.

The simplest model that accounts well for the spectra of both the total and the interferometric flux in our measurements requires only two gaussian components of different temperature, size and foreground dust absorption τ_{SiO} (see parameters in Table 1 and sketch in Fig. 1c). The first is a hot ($T > 800$ K) compact component. Its length along the source axis is resolved, and well constrained to 0.7 pc. The interferometric measurement sets only an upper limit (≤ 1 pc) to its orthogonal width. The second is a warm, well resolved component with best-fit temperature $T = 320$ K. Its size is well constrained by our observations with two baselines.

The silicate absorption feature around 10 μm is much more pronounced in the interferometric spectra (Fig. 2b, c) than in the total spectrum (Fig. 2a). In our best-fitting model, we find an spatially averaged peak silicate absorption depth, $\langle\tau_{\text{SiO}}\rangle$, of 0.3 in front of the extended $T = 320$ K component plus an additional $\langle\tau_{\text{SiO}}\rangle = 1.8$ in front of the hot component. Thus, the hot component seems to be embedded within the warm one. Presumably, much of the absorption towards the hot component actually occurs in the inner regions of the warm extended component. A ray from the hot dust, penetrating the full thickness of the larger, outer component, will suffer more absorption than the radiation from cooler dust in the outer regions of this component.

The profile of the silicate absorption towards the hot component does not fit well to the profiles of common olivine-type silicate dust, because those begin to drop already at $\sim 8 \mu\text{m}$, while the interferometric spectrum with the highest spatial resolution (Fig. 2c) does not seem to be affected at wavelengths $< 9 \mu\text{m}$. We obtain a much better fit by using the profile of calcium aluminium silicate ($\text{Ca}_2\text{Al}_2\text{SiO}_7$), a high-temperature dust species found in some supergiant stars¹². Although radiation transfer effects might fill in some of the dip near 8 μm even with ‘normal’ silicates, our modelling of this effect does not produce acceptable fits to the data. Other non-olivine dust compositions, however, cannot currently be excluded. In any case, our observations require special dust properties in the innermost 2 pc around the nucleus of NGC 1068. Indications of non-standard dust properties have also been reported for other Seyfert 2 galaxies¹³.

Table 1 Dust emission components in the nucleus of NGC 1068

Component	T (K)	$\Delta \parallel \text{jet}$ (mas)	$\Delta \perp \text{jet}$ (pc)	τ_{SiO}
Hot	> 800	10 ± 2	0.7 ± 0.2	2.1 ± 0.5
Warm	320 ± 30	30 ± 5	2.1 ± 0.4	0.3 ± 0.2

Each component is taken to be a two-dimensional gaussian aligned with one axis parallel to the radio jet. Sizes Δ are given as full-width at half-maximum of gaussian profiles. We have assumed emission from each component with a blackbody spectrum of the given temperature with constant emissivity. Emissivity decreasing rapidly with increasing wavelength ($\sim \lambda^{-1}$ or steeper), as might be expected from optically thin dust emission, is not consistent with the data. In front of this emission we impose an absorption of $\exp(-\tau(\lambda))$. $\tau(\lambda)$ varies with λ to match known silicate profiles as described in the text. The peak value of $\tau(\lambda)$ for each component is given here as τ_{SiO} . The optical depth τ in front of the hot component is in fact the sum of the contributions of the hot and warm components, because the model assumes that the absorption in the warm component also reduces the hot component. The errors refer to the full range of uncertainty when varying all parameters simultaneously. We assume the standard distance²¹ of 14.4 Mpc (that is, 1 pc = 14.3 mas). The temperature of the hot component is a lower bound; the blackbody profiles from all temperatures above 800 K are indistinguishable within the observed wavelength band.

This simple heuristic model describes the basic observed scale sizes of the dust emission in NGC 1068. But how does it relate to the true physical distribution of dust in active nuclei? The generic properties of dusty tori in AGNs have been discussed³. Because previously no observational means were available to resolve the dust distribution, early models for their dust emission^{14–16} aimed mainly to reproduce the overall spectral energy distribution of Seyfert galaxies and quasars. These were calculated by two-dimensional radiative transfer through a smooth dust distribution. They can be divided in two classes: compact tori¹⁴, where the dust fills a thick cylinder (with an axial hole) of a few parsecs in diameter, and extended fat disk models^{15,16}, with diameters of tens to hundreds of parsecs. Common to both classes was a very high absorption optical depth ($A_V > 100$, that is, $\tau_{12\mu} > 4$) in the equatorial plane. In preparation for our observations, we investigated the expected dust emission maps from those models by using three-dimensional radiative transfer¹⁷. They show that hot dust can only be observed from those surfaces that are both freely exposed to the radiation from the accretion disk and unobscured along the line of sight from the observer. In consequence, the observable structures at 8 and 12 μm are almost identical, and suffer very similar dust absorption. Obviously, this is not what we observe in NGC 1068.

The unification between Seyfert 1 and 2 galaxies¹ already seems to demand a geometrically thick dust distribution, the vertical height h (above the midplane) of which should be similar to its radius r , that is $h/r \approx 1$. By resolving the heated dust structure in NGC 1068 with MIDI/VLTI, we have demonstrated now that indeed $h/r \approx 0.6$, even when allowing for projection effects. (Such effects seem to be small, because the axis of NGC 1068 as given by the outflow cone of ionized gas is inclined by only $\sim 5^\circ$ out of the plane of sky¹⁸, thus excluding that an inclined thin disk could explain the apparent h/r).

Moreover, this thick structure is located at $r \leq 2\text{ pc}$. Thus, irradiation by the nuclear source will produce a hot inner wall—a ‘funnel’, which we identify as our hot component. As the average cloud temperatures are low, $T < 1,000\text{ K}$, it is not possible for gas pressure to support this structure vertically against gravity in the nuclear potential of NGC 1068.

In reality, it is very unlikely that gas and dust are distributed homogeneously in the torus: they will rather be confined to individual clouds³. A recent model for the overall spectral energy distributions¹⁹ argues for a volume filling factor of $< 10\%$, and typically 5–10 clouds along an equatorial line of sight towards the centre. Much better resolution ($< 1\text{ mas}$) would be required to distinguish such a torus from a homogeneous one. Turbulent motions of these clouds with average velocities $\langle v_T \rangle \approx 100\text{ km s}^{-1}$ (that is, similar to the random velocities that support the nuclear star cluster) could prevent the cloud system from collapsing. However, given these velocities and the large number of clouds along any line of sight, collisions between those clouds would be frequent, supersonic and highly inelastic. The turbulent motion would be damped within about one orbital period, $t_{\text{orb}}(at\ r = 2\text{ pc}) \approx 10^5\text{ yr}$. Thus, a continuous injection of kinetic energy into the cloud system seems to be required. To our knowledge, none of the current models of AGNs provide a convincing solution to this problem, although a very young nuclear star cluster might be able to provide the energy (but see ref. 3). This explanation requires AGN activity and nuclear starbursts to be intimately connected. So the present high-resolution observations of the nucleus of an active galaxy by infrared interferometry demand a better understanding of the physics of these spectacular objects. □

Received 3 February; accepted 1 April 2004; doi:10.1038/nature02531.

1. Antonucci, R. Unified models for active nuclei and quasars. *Annu. Rev. Astron. Astrophys.* **31**, 473–521 (1993).
2. Urry, C. M. & Padovani, P. Unified schemes for radio-loud active galactic nuclei. *Publ. Astron. Soc. Pacif.* **107**, 803–845 (1995).
3. Krolik, J. H. & Begelman, M. C. Molecular tori in Seyfert galaxies: Feeding the monster and hiding it. *Astrophys. J.* **329**, 702–711 (1998).

4. Antonucci, R. J. & Miller, J. S. Spectropolarimetry and the nature of NGC 1068. *Astrophys. J.* **297**, 621–632 (1985).
5. Lumsden, S. L. et al. Near- and mid-infrared imaging polarimetry of NGC 1068. *Mon. Not. R. Astron. Soc.* **303**, 209–220 (1999).
6. Leinert, C. et al. Ten-micron instrument MIDI: getting ready for observations on the VLTI. *Proc. SPIE* **4838**, 893–904 (2003).
7. Glindermann, A. et al. The VLTI—A Status Report. *Proc. SPIE* **4838**, 89–100 (2003).
8. Muxlow, T. W. B., Pedlar, A., Holloway, A. J., Gallimore, J. F. & Antonucci, R. R. J. The compact radio nucleus of the Seyfert galaxy NGC 1068. *Mon. Not. R. Astron. Soc.* **278**, 854–860 (1996).
9. Bock, J. J. et al. High spatial resolution imaging of NGC 1068 in the mid-infrared. *Astron. J.* **190**, 2904–2919 (2000).
10. Lutz, D. et al. ISO-SWS spectroscopy of NGC 1068. *Astrophys. J.* **536**, 697–709 (2000).
11. Kemper, F., Vriend, W. J. & Tielens, A. G. G. M. ISO spectra of Sgr A. *Astrophys. J.* (submitted).
12. Speck, A. K., Barlow, J. J., Sylvester, F. J. & Hofmeister, A. M. Dust features in the 10 micron infrared spectra of oxygen-rich evolved stars. *Astron. Astrophys. Suppl.* **146**, 437–464 (2000).
13. Maiolino, R. et al. Dust in active nuclei. I. Evidence for anomalous properties. *Astron. Astrophys.* **365**, 28–36 (2001).
14. Pier, E. A. & Krolik, J. H. Infrared spectra of obscuring dust tori around active galactic nuclei. *Astrophys. J.* **401**, 99–109 (1992).
15. Efstathiou, A., Hough, J. H. & Young, S. A model for the infrared continuum spectrum of NGC 1068. *Mon. Not. R. Astron. Soc.* **277**, 1134–1144 (1995).
16. Granato, G. L., Danese, L. & Franceschini, A. Thick tori around active galactic nuclei: the case for extended tori and consequences for their X-ray and infrared emission. *Astrophys. J.* **486**, 147–159 (1997).
17. Schartmann, M. *Dust Tori in Active Galactic Nuclei*. Thesis, Univ. Heidelberg (2003).
18. Crenshaw, D. M. & Kraemer, S. B. Resolved spectroscopy of the narrow-line region in NGC 1068: Kinematics of the ionized gas. *Astrophys. J.* **532**, L101–L104 (2000).
19. Nenkova, M., Ivezić, Z. & Elitzur, M. Dust emission from active galactic nuclei. *Astrophys. J.* **570**, L9–L12 (2002).
20. Bruhweiler, F. C., Miskey, C. L., Smith, A. M., Landsman, W. & Malumuth, E. Ionization, extinction and spiral structure in the inner disk of NGC 1068. *Astrophys. J.* **546**, 866–876 (2001).
21. Bland-Hawthorn, J. et al. The Ringberg standards for NGC 1068. *Astrophys. Space Sci.* **248**, 9–19 (1997).

Acknowledgements This Letter is based on observations carried out at the European Southern Observatory, Paranal, Chile, in the framework of the VLTI science demonstration program. MIDI could not have been built without contributions from many people, including E. J. Bakker, P. Ballester, S. Beckwith, P. Biereichel, A. Böhm, W. D. Cotton, S. Damstra, A. Glindemann, B. Grimm, J. de Jong, N. Haddad, H. Hanenburg, T. Henning, N. Housen, S. Hippler, W. Laun, S. Ligor, R. Lenzen, B. Lopez, O. von der Lühe, R. J. Mathar, J. Meisner, S. Morel, W. Morf, U. Neumann, R.-R. Rohloff, P. Schuller, N. Salm, C. Storz, A. Wallander and K. Wagner. We thank ESO staff at Garching, Santiago de Chile, and Cerro Paranal (Chile), for operating MIDI.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.J. (jaffe@strw.leidenuniv.nl).

Dislocation-driven surface dynamics on solids

S. Kodambaka, S. V. Khare, W. Świąch, K. Ohmori, I. Petrov & J. E. Greene

Frederick Seitz Materials Research Laboratory and the Department of Materials Science, University of Illinois, 104 South Goodwin Avenue, Urbana, Illinois 61801, USA

Dislocations¹ are line defects that bound plastically deformed regions in crystalline solids. Dislocations terminating on the surface of materials can strongly influence nanostructural and interfacial stability, mechanical properties, chemical reactions, transport phenomena, and other surface processes. While most theoretical and experimental studies have focused on dislocation motion in bulk solids under applied stress^{2,3} and step formation due to dislocations at surfaces during crystal growth^{4–7}, very little is known about the effects of dislocations on surface dynamics and morphological evolution. Here we investigate the near-equilibrium dynamics of surface-terminated dislocations using low-energy electron microscopy⁸. We observe, in real time, the thermally driven nucleation and shape-preserving growth of spiral steps rotating at constant temperature-dependent angular

velocities around cores of dislocations terminating on the (111) surface of TiN in the absence of applied external stress or net mass change. We attribute this phenomenon to point-defect migration from the bulk to the surface along dislocation lines. Our results demonstrate that dislocation-mediated surface roughening can occur even in the absence of deposition or evaporation, and provide fundamental insights into mechanisms controlling nanostructural stability.

Figure 1A and B illustrates the nucleation and growth of TiN(111) spiral steps and loops at the cores of dislocations terminating on TiN(111) surfaces, as observed *in situ* by low-energy electron microscopy (LEEM)⁹ during annealing in N₂. (We define a loop as a two-dimensional (2D) island formed around a surface step segment pinned at both ends by dislocations.) Fig. 1A shows the formation of a concentric spiral step structure around a dislocation core, while the images in Fig. 1B capture the nucleation and growth of a closed loop. The spiral and loop steps both exhibit three-fold symmetry with a truncated-hexagonal shape, the near-equilibrium shape of 2D TiN(111) islands¹⁰, indicative of fast step-edge diffusion. Note that these spirals and loops are observed during annealing with no net mass gain or loss and, as we will show, are quantitatively and qualitatively different from the growth structures predicted by Burton, Cabrera and Frank⁷ ('BCF'), and step curvature-driven surface dynamics¹¹.

A remarkable feature is revealed in Fig. 1A, a–d: the spiral steps rotate around the dislocation core, resulting in an increase in the total step length with time t as the spirals undergo a shape-preserving anticlockwise motion with a constant angular velocity ω . That is, the shape and size of the spiral are periodic with time $\tau = 2\pi/\omega$. For the spiral shown in Fig. 1A, a–d, $\tau = 47$ s corresponds to one complete rotation at the annealing temperature $T = 1,688$ K; the spiral shown in Fig. 1A, a, is geometrically identical to the one in Fig. 1A, d. During the period τ , the layer labelled 3 disappears from the field of view in Fig. 1A, c, and a new layer, labelled 0, is formed (Fig. 1A, d) as the dislocation core climbs by a unit step height, $a_{\perp} = 2.4$ Å, normal to the surface.

Nucleation and growth of a loop originating at a defect is shown in Fig. 1B, a–c. Upon detaching from the defect, the expanding loop

(Fig. 1B, d) regains the equilibrium shape of 2D TiN(111) islands¹⁰. This process, like the nucleation and growth of spiral steps, is also periodic. For the loops shown in Fig. 1B, a–d, $\tau = 85 \pm 5$ s corresponds to the loop nucleation period at $T = 1,653$ K.

Analogous to the growth spirals and loops observed in bulk solids owing to the operation of Frank–Read¹² and/or Bardeen–Herring¹³ sources under applied stress¹⁴, and on surfaces during crystal growth as predicted by the BCF theory⁷, the features observed in Fig. 1A and B are due to surface steps pinned by a single dislocation (in the case of a spiral) and a pair of oppositely signed dislocations (for a loop), respectively. For a single dislocation, the steps emanating from the cores wind into expanding spirals. With a pair of oppositely signed dislocations, as illustrated in the schematic diagram of Fig. 1C, the expanding spirals originating from the cores rotate in opposing directions, eventually coming into contact to form a loop that increases in size.

Figure 2a is a representative plot of the areas A of a TiN(111) spiral and an adjacent surface loop (both are shown in the corresponding LEEM image, right) as a function of time t at 1,653 K. The two sets of data correspond to successive spirals and loops originating periodically from their respective sources. We observe that at constant temperature, spirals and loops have the same generation times ($\tau = 85 \pm 5$ s for the examples shown in the LEEM image in Fig. 2a, $T = 1,653$ K) and that their areas increase linearly with time. The similarity in dA/dt and ω for the two types of structures suggests that the same mechanism controls their growth. The observed step growth is a localized process that occurs only at the dislocation cores. This is in contrast to the BCF model, where the

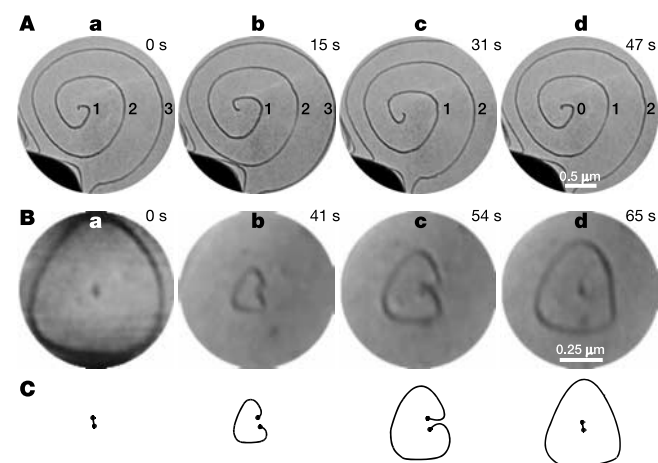


Figure 1 Nucleation and growth of spiral steps on TiN(111). Low-energy electron microscopy (LEEM) images showing nucleation and growth of bilayer-height surface steps at the cores of dislocations terminating on TiN(111) terraces during annealing in N₂ at temperatures T . The images were acquired as a function of time t . The dark lines in the LEEM images are <110>-oriented bilayer-height steps (the [111] direction in B1-NaCl structure TiN is polar, consisting of alternating layers of Ti and N atoms). **A**, Field of view ~ 2.53 μm ; $T = 1,688$ K. Spiral steps form owing to the pinning of a step edge at a dislocation core. **B**, Field of view ~ 0.93 μm ; $T = 1,653$ K. Loops nucleate and grow centred around a step edge pinned at both ends by dislocations of opposite sign. **C**, Schematic diagram of the loop generation process observed in **B**.

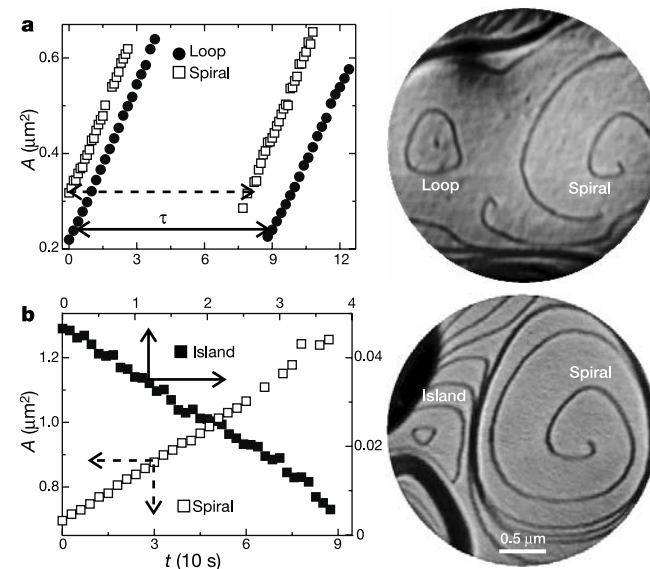


Figure 2 Area versus annealing time for 2D TiN loops, spirals and islands on TiN(111). **a**, Time-dependent areas A of 2D TiN loops and spirals on a TiN(111) surface (shown in the image to the right) during annealing at $T = 1,653$ K. (We only measure areas of loops that formed upon detaching from the central core and have attained near-equilibrium truncated-hexagonal shapes as in Fig. 1B, d.) Spiral areas are calculated as $A(t) = \frac{1}{2} \int_{\theta_1}^{\theta_2} [r(\theta, t)]^2 d\theta$. The spiral shape function $r(\theta, t)$ is defined for θ in the range 0 to ∞ , and $\theta_1 \leq \theta \leq \theta_2$ corresponds to an outwardly moving spiral step segment far from the core. Because loop areas increase linearly with time, while the growth rates of spirals with high curvatures ($> 2 \times 10^{-4} \text{ Å}^{-1}$) near the core are nonlinear, we only measure areas of spirals corresponding to linear growth rates with $\theta_1 = \pi$ and $\theta_2 = 3\pi$. τ is the average time required to generate successive spirals and loops that have the same area. **b**, The time-dependent areas A of spirals and 2D TiN islands on TiN(111) terraces (shown in the image to the right) during annealing at $T = 1,694$ K. Note that the left and bottom axes of the plot correspond to that of the spiral, while the right and top axes correspond to the island. The spiral area increases linearly with time while the island area decreases. The fields of view in both images are ~ 2.75 μm .

growth flux is non-localized, resulting in a curvature-dependent step velocity of spirals and loops whose areas increase nonlinearly with time⁷.

Figure 2b contains A versus t plots for a 2D TiN(111) adatom island and an adjacent spiral (shown in the associated LEEM image, right) during annealing at 1,694 K. Note that the island area decreases linearly with t while the spiral area increases. Previous studies¹⁵ have shown that step-edge attachment/detachment is the rate-limiting mechanism for the curvature-driven decay of 2D TiN adatom and vacancy islands on TiN(111) terraces. The above results, typical of data from over 50 adatom islands and 9 spirals acquired at temperatures 1,500–1,700 K, demonstrate that the mechanism of surface spiral and loop growth is quite distinct from that of 2D TiN(111) island decay (Ostwald ripening)¹¹. We suggest, and provide experimental evidence below, that the observed nucleation and growth of spiral steps during annealing is due to dislocation-assisted bulk diffusion.

Our model for this phenomenon is based upon two assumptions. (1) A non-equilibrium concentration of point defects exists in the bulk. This is reasonable based upon the fact that TiN is known to have a very wide single-phase region and can sustain both high anion (N) and cation (Ti) vacancy concentrations¹⁶. Given that we observe steps emanating from the grooves (the thick dark lines visible in the LEEM images in Figs 1, 2 and 3) annihilating the spiral steps, we conclude that spirals grow inward, normal to the surface.

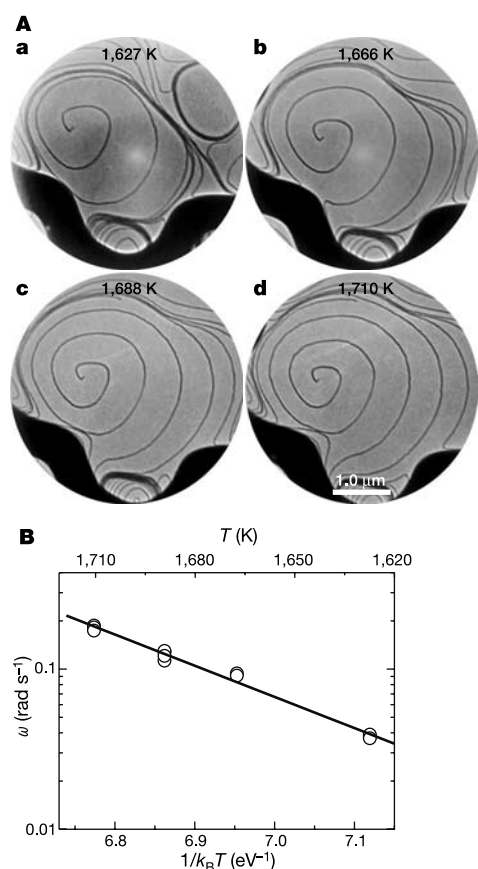


Figure 3 Temperature dependence of angular velocities of spirals on TiN(111). **A**, LEEM images (fields of view $\sim 3.79 \mu\text{m}$) of TiN spiral steps on a TiN(111) terrace. The images were acquired as a function of temperature T . **B**, Angular velocity ω of the spirals, plotted as a function of $1/T$. ω is determined by fitting a function of the form $r(\theta, t) = a(\theta)[\theta + \omega t]$, where $a(\theta)$ is a time-independent shape function, to the time-dependent spiral step boundary coordinates obtained from the LEEM images. The solid line is fitted to the $\omega(T)$ data with a function of the form $\omega(T) = \omega_0 \exp(-E_d/kT)$ using least-squares analyses. We obtain $\omega_0 = 10^{12 \pm 0.5} \text{ s}^{-1}$ with $E_d = 4.5 \pm 0.2 \text{ eV}$.

Note, however, that there is no net mass gain or loss. The surface point-defect concentration C_s^{eq} , however, is at thermal equilibrium, because a N-terminated surface is energetically favourable for TiN(111) (ref. 17). (2) Dislocation cores emit/absorb point defects at a thermally activated time-independent rate R ; a plausible assumption given that dislocation cores act as sources/sinks for point defects^{18,19}.

Consider for simplicity a circular loop of radius r_{loop} centred around a core region of finite radius r_{core} . The surface point-defect concentration $C(r)$ at any position r , with $r_{\text{core}} < r < r_{\text{loop}}$, is given by the diffusion equation $\partial C(r)/\partial t = D_s \nabla^2 C(r)$, where D_s is the surface diffusivity. Within the quasistatic approximation in which surface diffusion of adatoms is much faster than the step-edge velocity, $\partial C(r)/\partial t = 0$. The general solution of the resulting Laplace equation $\nabla^2 C(r) = 0$ is $C(r) = C_1 \ln(r) + C_2$, where the constants C_1 and C_2 are determined by the boundary conditions at the core and the loop. From assumption (2) above, the flux $-D_s \nabla C(r)$ of point defects at the core is

$$-D_s \nabla C(r)|_{r=r_{\text{core}}} = \frac{R}{2\pi r_{\text{core}}} \quad (1a)$$

while at the loop:

$$-D_s \nabla C(r)|_{r=r_{\text{loop}}} = k_s [C(r_{\text{loop}}) - C_{\text{loop}}^{\text{eq}}] \quad (1b)$$

In equation (1b), k_s is the rate of attachment/detachment at the step and $C_{\text{loop}}^{\text{eq}} = C_s^{\text{eq}} \exp(\mu_{\text{loop}}/k_B T)$, from the Gibbs–Thomson relation¹¹, is the equilibrium point-defect concentration due to the curvature-dependent chemical potential μ_{loop} associated with the loop. Solving for C_1 and C_2 yields the normal component of the loop velocity dr_{loop}/dt as

$$\frac{dr_{\text{loop}}}{dt} = \Omega k_s [C(r_{\text{loop}}) - C_{\text{loop}}^{\text{eq}}] = \frac{\Omega}{2\pi r_{\text{loop}}} R \quad (2)$$

where Ω is the unit TiN molecular area. In deriving the above formalism, we have neglected the curvature-driven flux from the loop to the local environment. In the detachment-limited regime, the contribution to dr_{loop}/dt due to this flux, which is proportional to $-1/r_{\text{loop}}$, is small and hence will have little effect on the growth velocity dr_{loop}/dt in equation (2). The important point is that the form of equation (2) is qualitatively different from the parallel equation describing the 2D island decay (Ostwald ripening) process¹⁵. We note that equation (2) is also valid, without loss of generality, for non-circular loops and spiral steps far from the core. Clearly, from equation (2), the loop (and spiral) growth rate $dA/dt = \Omega R$ is time-invariant, consistent with the constant slopes obtained in our measurements of A versus t in Fig. 2.

For the steady-state growth of a spiral rotating around a dislocation core, we obtain from mass conservation the relation $\omega = 2\pi \Omega R/A_0$, indicating that the frequency ω is thermally activated. A_0 in the above relation is the area encompassing the spiral structure over which R is non-negligible. Substituting the experimental data for dA/dt and ω into our model, we obtain A_0 values between 0.25 and $0.32 \mu\text{m}^2$, which are less than the areas encompassed by spiral boundaries. Our measurements of $(dA/dt)/\omega$ (which is proportional to A_0) versus T suggest that A_0 is independent of temperature. However, understanding the physical significance of A_0 requires further investigation.

Typical ω values, determined from separate sets of LEEM images acquired at four different temperatures T (see Fig. 3A) are plotted as a function of $1/T$ in Fig. 3B. From the $\omega(T)$ data, we find an activation energy $E_d = 4.5 \pm 0.2 \text{ eV}$, with a prefactor of $10^{12 \pm 0.5} \text{ s}^{-1}$, compared to $2.3 \pm 0.6 \text{ eV}$ for the decay of 2D TiN islands on TiN(111) (ref. 15). A constant ω for a given T implies that the dislocation climb velocity normal to the surface is also time-invariant. Hence, E_d corresponds to the activation barrier for dislocation motion, which is generally associated with point-defect formation and migration along the dislocation (also referred to as

pipe “diffusion”²⁰. The fact that we measure a higher activation barrier E_d for spiral growth than for island decay (and E_d is significantly lower than the desorption energies for TiN and Ti adspecies¹⁷) provides strong evidence that the dominant mass transport is along dislocation lines rather than surface diffusion or evaporation. This is physically reasonable because bulk point-defect migration, expected to influence surface dynamics at sufficiently high temperatures²¹, has a smaller activation barrier along dislocation lines²⁰ than in dislocation-free areas.

In formulating equation (2), we assume an equilibrium defect concentration $C_{\text{loop}}^{\text{eq}}$ associated with the loop. This is justified for the first loop, as steps near the core and far from the boundary maintain a truncated-hexagonal shape with three-fold symmetry, the equilibrium shape of TiN(111) steps. (This observation also suggests that localized growth flux at the cores may have negligible effect on step shapes.) However, far from the cores we find, first, steps with non-equilibrium shapes that vary with the spiral geometry and, second, step bunching. We attribute these observations to the presence of grooves bounding the spirals. Hence, modelling growth kinetics of multiple loop/spiral steps requires a better understanding of the effects of the geometric constraints imposed by physical boundaries on step chemical potentials.

Finally, our model assumes a constant rate R , which depends on two driving forces: first, equilibration of bulk point-defects, and second, minimization of dislocation line energy. In the first case, R is a function of the concentration gradient between the surface and the bulk, which decreases with time (K. McCarty and N. C. Bartelt, personal communication). Hence, this driving force terminates upon equilibration of the bulk point-defect concentration. In the second case, however, spiral and loop growth will continue to occur until the disappearance of the dislocation core. Our preliminary A versus t data for loops and spirals indicate that dA/dt systematically decreases by $\leq 4\%$ for successive generation of new spirals and loops. This, however, is within experimental uncertainties of $\sim 10\%$. Thus, additional experiments for extended annealing times are necessary to quantitatively determine the time-dependence, if any, in dA/dt and hence the dominant driving force.

We expect the near-equilibrium spiral step growth process described above to be general, and that it will be observed for other materials. In contrast to recent work showing surface^{15,22} and bulk²¹ transport dominated 2D island dynamics, the present study demonstrates that facile bulk diffusion along the dislocations can lead to both qualitatively and quantitatively different surface evolution kinetics. □

Received 10 October 2003; accepted 11 March 2004; doi:10.1038/nature02495.

- Cottrell, A. H. in *Dislocations and Plastic Flow in Crystals* (eds Fowler, R. H., Kapitza, P., Mott, N. F. & Bullard, E. C.) Ch. 1 (Oxford Univ. Press, London, 1953).
- Bulatov, V., Abraham, F. F., Kubin, L., Devincere, B. & Yip, S. Connecting atomistic and mesoscale simulations of crystal plasticity. *Nature* **391**, 669–672 (1998).
- Lagow, B. W. et al. Observation of dislocation dynamics in the electron microscope. *Mater. Sci. Eng. A* **309–310**, 445–450 (2001).
- Verma, A. R. & Amelinckx, S. Spiral growth on carborundum crystal faces. *Nature* **167**, 939–940 (1951).
- Teng, H. H., Dove, P. M., Orme, C. A. & De Yoreo, J. J. Thermodynamics of calcite growth: Baseline for understanding biomineral formation. *Science* **282**, 724–727 (1998).
- Pina, C. M., Becker, U., Risthaus, P., Bosbach, D. & Putnis, A. Molecular-scale mechanisms of crystal growth in barite. *Nature* **395**, 483–486 (1998).
- Burton, W. K., Cabrera, N. & Frank, F. C. The growth of crystals and the equilibrium structure of their surfaces. *Phil. Trans. R. Soc. Lond. A* **243**, 299–358 (1951).
- Bauer, E. Low-energy electron microscopy. *Rep. Prog. Phys.* **57**, 895–938 (1994).
- Tromp, R. M. & Reuter, M. C. Design of a new photo-emission/low-energy electron microscope for surface studies. *Ultramicroscopy* **36**, 99–106 (1991).
- Kodambaka, S. et al. Absolute TiN(111) step energies from analysis of anisotropic island shape fluctuations. *Phys. Rev. Lett.* **88**, 146101 (2002).
- Zinke-Allmang, M., Feldman, L. C. & Grabow, M. H. Clustering on surfaces. *Surf. Sci. Rep.* **16**, 377–463 (1992).
- Frank, F. C. & Read, W. T. Multiplication processes for slow moving dislocations. *Phys. Rev.* **79**, 722–723 (1950).
- Bardeen, J. & Herring, C. in *Imperfections in Nearly Perfect Crystals* (eds Shockley, W., Hollomon, J. H., Maurer, R. & Seitz, F.) Chap. 10 (Wiley, New York, 1952).
- Hull, D. & Bacon, D. J. *Introduction to Dislocations* Ch. 8 (Pergamon, Oxford, 1984).

- Kodambaka, S. et al. Size-dependent detachment-limited decay kinetics of two-dimensional TiN islands on TiN(111). *Phys. Rev. Lett.* **89**, 176102 (2002).
- Sundgren, J. E., Johansson, B. O., Rockett, A., Barnett, S. A. & Greene, J. E. in *Physics and Chemistry of Protective Coatings* (eds Greene, J. E., Sproul, W. D. & Thornton, J. A.) 95–115 (American Institute of Physics, New York, 1986).
- Gall, D., Kodambaka, S., Wall, M. A., Petrov, I. & Greene, J. E. Pathways of atomistic processes on TiN(001) and (111) surfaces during film growth: an *ab initio* study. *J. Appl. Phys.* **93**, 9086–9094 (2003).
- Weertman, J. & Weertman, J. R. in *Elementary Dislocation Theory* (eds M. E., Weertman, J. & Weertman, J. R.) Ch. 6 (Macmillan, New York, 1964).
- Justo, J. F., de Koning, M., Cai, W. & Bulatov, V. V. Point defect interaction with dislocations in silicon. *Mater. Sci. Eng. A* **309–310**, 129–132 (2001).
- Indenbom, V. L. & Saralidze, Z. K. in *Elastic Strain Fields and Dislocation Mobility* (eds Indenbom, V. L. & Lothe, J.) Ch. 10 (Elsevier Science, Amsterdam, 1992).
- McCarty, K. E., Nobel, J. A. & Bartelt, N. C. Vacancies in solids and the stability of surface morphology. *Nature* **412**, 622–625 (2001).
- Giesen, M. Step and island dynamics at solid/vacuum and solid/liquid interfaces. *Prog. Surf. Sci.* **68**, 1–153 (2001).

Acknowledgements This work was supported by the US Department of Energy (DOE), Division of Materials Science, through the University of Illinois Frederick Seitz Materials Research Laboratory (FS-MRL). We thank C. P. Flynn, H. Birnbaum, R. J. Pflueger and P. O. Å. Persson for discussions and critical reading of the manuscript. We also appreciate the use of the facilities in the Center for Microanalysis of Materials, partially supported by DOE, at the FS-MRL.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.K. (kodambak@uiuc.edu).

Polymerization within a molecular-scale stereoregular template

Takeshi Serizawa*, Ken-ichi Hamada & Mitsuru Akashi*

Department of Nanostructured and Advanced Materials, Graduate School of Science and Engineering, Kagoshima University, 1-21-40 Korimoto, Kagoshima 890-0065, Japan

* Present addresses: Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan (T.S.); Department of Molecular Chemistry, Graduate School of Engineering, Osaka University, 2-1 Yamada-oka, Suita 565-0871, Japan (M.A.)

Enzymes efficiently synthesize biopolymers by organizing monomer units within regularly structured molecular-scale spaces and exploiting weak non-covalent interactions, such as hydrogen bonds, to control the polymerization¹ process. This ‘template’ approach is both attractive and challenging for synthetic polymer synthesis, where structurally regulated molecular-scale spaces could in principle provide solid-phase reaction sites for precision polymerization. Previously, free-radical polymerization of methyl methacrylate in solutions containing stereoregular isotactic (it) or syndiotactic (st) poly(methyl methacrylate) (PMMA) has been shown to result in template synthesis^{2,3} of the opposite PMMA based on stereocomplex formation^{4,5} with van der Waals interactions. However, using the structure of a solid to determine the stereochemical structure of a polymer has not been satisfactorily achieved⁶. Here we show that macromolecularly porous ultrathin films, fabricated by a single assembly step, can be used for the highly efficient stereoregular template polymerization of methacrylates through stereocomplex formation. This reaction mould accurately transfers its structural properties of stereoregularity, molecular weight and organization within the template to the new polymer.

Structurally regulated molecular-scale spaces (nanospaces) in ultrathin polymer films are potentially valuable as solid-phase reaction sites for precision polymerization. As the macromolecular mobility is restricted in the solid state, the template effect of a constituent polymer is potentially accelerated. It is, however, difficult

to prepare such films using conventional methods. We selected a methodology involving the solvent extraction of certain polymer components from ultrathin polymer films for the construction of porous host films for template polymerization. The precursor films were fabricated using layer-by-layer assembly⁷, which can be demonstrated by the simple alternate immersion of solid substrates into interactive polymer solutions. Although layer-by-layer assembly conventionally produces polyelectrolyte multilayers based on polyon complexes, stereocomplexes between it-PMMA and st-polymers of methacrylates were also assembled on solid substrates^{8–10}, thus forming ultrathin stereocomplex films with a double-stranded helical structure, in which it-PMMA was surrounded by twice the unit-molar amount of st-polymers (1:2 stoichiometry)⁵. In the case of the stereocomplex film composed of it-PMMA and st-PMAA, st-PMAA was selectively extracted from the film in an aqueous alkaline solution, resulting in a macromolecularly porous it-PMMA film¹¹. This porous film subsequently incorporated st-PMAA previously extracted, dependent on the st-PMAA concentration, thus demonstrating an unusual macromolecular recognition. However, it-PMMA was not extracted from the film in organic solvents, indicating that a versatile host composed of it- or st-polymers for template polymerization cannot be prepared by this method.

In this study, we focused on the formation of a stereocomplex between it-PMMA and st-PMAA with a 1:1 unit-molar stoichiometry^{12,13}. Layer-by-layer assembly of these polymers under suitable conditions produced a stereocomplex film with a 1:1 stoichiometry, and the subsequent selective extraction of a single component successfully produced porous st-PMAA or it-PMMA films, of which the macromolecular nanospaces could be used in stereoregular template polymerization of it-PMMA and st-PMAA, respectively. Although it is more difficult to obtain it-polymers of methacrylates by free-radical polymerization owing to the steric hindrance of lateral groups, the stereoregularity of polymers synthesized by this method was absolutely dependent on that of the template polymers. Accordingly, the present study demonstrated the novel template synthesis of valuable polymers using both weaker van der Waals interactions and artificially constructed nanospaces in films. The present template polymerization method is schematically represented in Fig. 1.

A 9-MHz quartz crystal microbalance (QCM) substrate^{8–11}, which could determine the amount of assembled polymer from its frequency decrease, was used to quantify the assembly of it-PMMA and st-PMAA. To obtain an assembly with 1:1 stoichiometry, dimethylformamide (DMF) and DMF/water (2/3, v/v) were selected as solvents for it-PMMA and st-PMAA, respectively. This solvent combination was determined by a detailed assembly

analysis, following methods used in a previous study¹³. Alternate immersion in both solutions produced a film with the objective stoichiometry of 1.0 ± 0.2 (st/it-unit ratio). The thickness and the mean roughness (R_a) for 6 cycles ((it-PMMA/st-PMAA)₆) were estimated to be 20.1 ± 0.3 nm and 3.2 nm, respectively. Considering the assembly amount analysed by the QCM ($4.1 \pm 0.1 \mu\text{g cm}^{-2}$), the mean density was estimated to be 1.0 g cm^{-3} . The immersion of the film in both chloroform and a 10 mM sodium hydroxide aqueous solution resulted in desorption of approximately half the amounts of polymers in the film. As the desorption amounts were consistent with the amounts of each polymer totally assembled, this suggests that it-PMMA and st-PMAA were selectively extracted from the film. Attenuated total reflection (ATR) spectra in the carbonyl vibration region also demonstrated the selective disappearance of peaks at $1,731 \text{ cm}^{-1}$ and $1,717 \text{ cm}^{-1}$, corresponding to respective it-PMMA and st-PMAA. The thickness after the extraction of it-PMMA and st-PMAA was 18.3 ± 2.8 nm and 19.4 ± 2.8 nm, respectively, and these values were significantly consistent with the initial thickness of the stereocomplex film within experimental error. In addition, the R_a was 2.1 nm and 2.8 nm, respectively, and was slightly smoother than before. These observations indicated that a porous film composed of st-PMAA or it-PMMA suitable for template polymerization was successfully prepared by the solvent extraction. It is difficult to explain reasonably why the films were not compressed by the

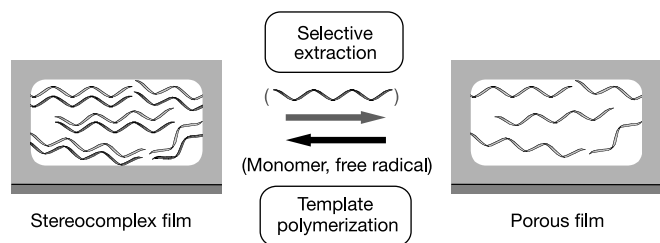


Figure 1 A schematic representation of template polymerization using ultrathin porous films. Layer-by-layer assembly prepares the ultrathin film of a stereocomplex comprising it-PMMA and st-PMAA. Two polymer chains are shown as rigid helical rods, but should be partially distorted and entangled in the film. There must be a disordered region of the stereocomplex. A single component is selectively extracted from the film, resulting in the preparation of the porous film with regular nanospaces. Then, the porous film is used as a reaction mould for free-radical template polymerization of MMA or MAA, followed by the regeneration of the stereocomplex film.

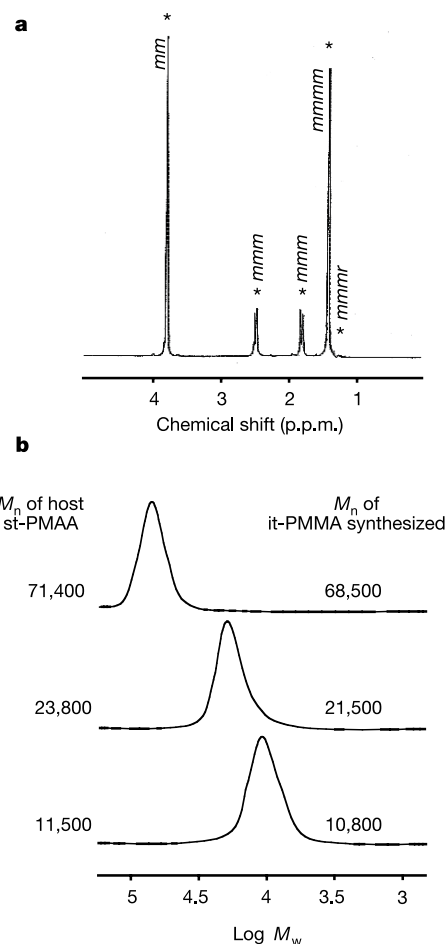


Figure 2 Characterization of it-PMMA polymerized in porous st-PMAA films for 3 h at 70 °C. **a**, A typical ^1H -NMR spectrum of it-PMMA polymerized in a porous st-PMAA (M_n 23,800) film. Symbols using m and r represent stereoregularity (see Methods). The asterisks indicate the peak position. **b**, SEC curves of it-PMMA polymerized in porous films of st-PMAAs with various molecular weights.

Table 1 Analytical data on stereoregular template polymerization of methacrylates

Host polymer	<i>mm:mr:rr</i>	M_n	M_w/M_n	Polymer synthesized	<i>mm:mr:rr</i>	M_n	M_w/M_n
st-PMAA	3:5:92	11500	1.7	it-PMMA	92:6:2	10800	2.5
st-PMAA	1:4:95	23800	1.3	it-PMMA	94:6:0	21500	2.0
st-PMAA	1:2:97	40500	1.5	it-PMMA	95:3:2	39300	1.9
st-PMAA	1:6:93	71400	1.6	it-PMMA	97:3:0	68500	2.4
it-PMMA	97:3:0	11100	1.2	st-PMMA	0:6:94	9800	2.0
it-PMMA	97:3:0	21050	1.4	st-PMMA	0:2:98	19200	2.1
it-PMMA	96:2:2	69700	1.2	st-PMMA	1:3:96	68900	2.0

Assemblies were demonstrated with it-PMMAAs ($mm > 96$, $M_n \approx 20,000$, $M_w/M_n < 2.0$) and st-PMAAs ($rr > 95$, $M_n \approx 40,000$, $M_w/M_n < 2.0$) for st-PMAA and it-PMMA hosts, respectively.

extraction. The terminals of the polymers as well as some segments that did not join in the complex might support the film structure. This observation was similar to the previously reported case of selective extraction from a double-stranded stereocomplex film¹¹.

For the first step of the analysis of template polymerization, the QCM substrate coated with the porous st-PMAA film was immersed in a 10 ml DMF solution of methyl methacrylate (MMA; 1.7 mg ml⁻¹) for 3 h at 70 °C in the presence of a free-radical initiator, 2,2'-azobisisobutyronitrile (5 mg ml⁻¹) under a nitrogen atmosphere. After gentle rinsing with DMF, approximately 90% of the it-PMMA previously extracted was recovered (the percentage was almost the same for all template st-PMAAs used). The ATR spectrum of the obtained film demonstrated the same peaks as those observed for the stereocomplex film, suggesting that it-PMMA was prepared, and that it formed a stereocomplex with the st-PMAA in the film. In order to analyse the stereoregularity and the molecular weight of the PMMA as the second step, the porous film was similarly prepared on silica particles (10 g, mean diameter 1.6 µm), following methods described in a previous study¹⁴. MMA was then polymerized, and the PMMA was characterized by extraction using chloroform. Approximately 0.4 g PMMA was obtained from the feeding of 1.5 g MMA per 500 ml DMF, thus resulting in a yield of approximately 25% using these polymerization conditions. A typical ¹H-NMR spectrum of the PMMA clearly demonstrated the it-specific polymerization of MMA in the porous st-PMAA film (Fig. 2a), thus indicating that the it-specific template polymerization was achieved on the basis of stereocomplex formation with the st-PMAA. Furthermore, size-exclusion chromatography (SEC) curves potentially demonstrated the control of molecular weights with a narrow distribution (Fig. 2b).

The experimental data from the template polymerization of MMA are summarized in Table 1. The isotacticities were greater than 92% in all cases. The number-average molecular weight (M_n) of the host st-PMAA significantly controlled the M_n of the it-PMMA, and was almost the same. These observations are surprising because the polymerization conditions, such as concentrations of monomers and initiators, were the same in all cases except for the molecular weight of the template st-PMAA. Although the same it-PMMA ($M_n \approx 20,000$) was used for the preparation of host films, the molecular weight of the it-PMMA prepared is absolutely dependent on that of the host st-PMAA. The properties of the template polymers corresponding to the 1:1 stereocomplex assembly were clearly transferred to the PMMAAs synthesized. The molecular weight distribution was small, compared with conventional free-radical polymerization. In fact, the it-PMMA recovered from the supernatant showed no stereoregularity ($mm:mr:rr = 21:68:11$) (see Methods) and a different molecular weight with a larger distribution (M_n 4,670; M_w/M_n 4.9) (M_w indicates the weight-average molecular weight). It is also notable that the it-PMMA polymerized was purified by a single precipitation in hexane. Accordingly, the control not only of the stereoregularity but also of the molecular weight of it-PMMA was realized by the present template polymerization method using macro-molecularly porous films. In solutions, it-specific template polymerization with 92% isotacticity in the presence of the template

st-PMMA has been demonstrated in cases of low conversions (<4%), low polymerization temperature (0 °C), and high molecular weights of templates ($\sim 10^6$) (ref. 6). Accordingly, the present polymerization of MMA was superior in both efficiency and performance to previous template polymerization.

The porous film of it-PMMA was used to prepare st-PMAA. Silica particles coated with the porous it-PMMA film were similarly used for free-radical polymerization of MAA in water in the presence of a water-soluble free-radical initiator, 2,2'-azobis(*N,N*-dimethyleisobutyramidine)dihydrochloride for 2 h at 40 °C under a nitrogen atmosphere. The PMAA prepared in the film was extracted in an alkaline solution, methylated to prepare PMMA, and then characterized, as also shown in Table 1. St-PMAA was successfully prepared by using the porous it-PMMA film, and the molecular weight of the host it-PMMA similarly regulated that of the st-PMAA. These observations indicate alternative utilization of the porous films for stereoregular template polymerization of methacrylates. We note that conventionally prepared spin- and solution-casting films did not realize the above template polymerization, even under the same polymerization conditions. Furthermore, template polymerization of the combination between it-PMMA and st-PMAA in the solution system only demonstrated the acceleration of polymerization rates, because the association is relatively weaker than that of stereoregular PMMAAs¹⁵. These observations also confirmed the potential of the present porous films. Preliminary experiments of template polymerization of MAA in a porous it-PMAA film¹¹ prepared from the 1:2 (it:st) complex assembly revealed that the molecular weight of the st-PMMA polymerized was approximately twice that of the template it-PMMA (K.-i.H., T.S. and M.A., manuscript in preparation), indicating the potential transferring of the assembly structure as well as further applications of porous thin films.

We now consider the polymerization mechanism in the films. Two contradictory polymerization processes would be expected: (1) polymerization after sufficient coordination of monomers with templates, and (2) stepwise polymerization with gradual incorporation (or slow diffusion) of monomers into the films. The quantitative QCM analysis did not show the sorption of monomers into the films, and did not support the former mechanism. In order to

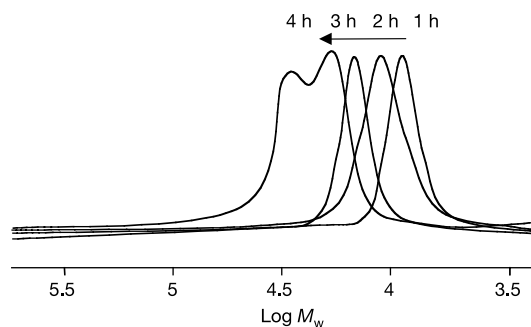


Figure 3 Time-dependent change in SEC curves of it-PMMA polymerized in a porous st-PMAA (M_n 40,500) film.

understand these mechanisms in detail, time-dependent SEC curves of the it-PMMA prepared by the st-PMAA host were obtained (Fig. 3). The curve with a single and narrow peak width shifted to a higher molecular weight, thereby supporting the latter mechanism, which is similar to living radical polymerization rather than conventional free-radical polymerization. We also suggest that the growing polymer chain migrates along the complementary template chain until the prepared nanospace is full. Although the curve was bimodal after 4 h, this is possibly due to the binding of terminal radicals in the film. As we stopped the polymerization of MMA after 3 h by opening the reaction vessel to the air, the radical terminals, which had been present in the films for at least 3 h, seemed to react with oxygen. The molecular weight of the st-PMMA prepared was similarly shifted, and a similar mechanism was supported.

This is, to our knowledge, the first report of an artificially constructed polymerization template and subsequent polymerization with the efficient transfer of structural information, similar to that seen in biosynthetic processes. Porous polymer matrices provided a reaction mould for stereoregular polymerization of methacrylates. It- and st-polymers with high stereoregularity and with a narrow molecular weight distribution were successfully prepared. Because particles coated with functional ultrathin films are readily handled and reusable (the porous films were reused at least 3 times), we expect large-scale synthesis of the stereoregular polymers to be achieved in the near future. Easier synthesis of stereoregular polymers of methacrylates potentiates their applications in technological and biomedical fields. Regulated nanospaces prepared in synthetic polymer assemblies should open the way to a new field of macromolecular recognition and synthesis. □

Methods

Assembly

The 9-MHz QCM with polished gold electrodes ($R_a = 1.8$ nm) was alternately immersed into the DMF solution of it-PMMA (1.7 mg ml^{-1}) and the DMF/water (2/3, v/v) solution of st-PMAA (1.5 mg ml^{-1}) for 5 min at 25°C . After each immersion step, the QCM was gently rinsed with the same solvent, dried with nitrogen gas, and the frequency shift then measured. The shift was converted to the assembly amount by applying Sauerbrey's equation^{8–11}. The assembly was initiated with it-PMMA in all cases.

Characterization

The QCM electrode with a refractive surface was used directly to observe the ATR spectra. The interferograms were co-added 50 times, and Fourier transformed at a resolution of 4 cm^{-1} . The film thickness was analysed by the scratching mode of an atomic force microscope (AFM). The roughness was also analysed by AFM. 400-MHz NMR (chloroform solvent) was used to analyse stereoregularity, in which *m* and *r* indicate the it and st-diads of meso and racemo, respectively. The triad (*mm*, *mr* and *rr*) and the tetrad (for example, *mmmm*) are represented similarly. SEC (THF solvent) was performed by PMMA standard.

Received 4 October 2003; accepted 29 March 2004; doi:10.1038/nature02525.

1. Alberts, B. *et al.* *Essential Cell Biology* 2nd edn (Garland, New York, 2003).
2. Buter, R., Tan, Y. Y. & Challa, G. Radical polymerization of methyl methacrylate in the presence of isotactic poly(methyl methacrylate). *J. Polym. Sci. A* **10**, 1031–1049 (1972).
3. Buter, R., Tan, Y. Y. & Challa, G. Radical polymerization of methyl methacrylate in the presence of stereoregular poly(methyl methacrylate). II. Syndiotactic PMMA as matrix. *J. Polym. Sci. Polym. Chem. Edn* **11**, 1003–1011 (1973).
4. Spěváček, J. & Schneider, R. Aggregation of stereoregular poly(methyl methacrylates). *Adv. Colloid Interface Sci.* **27**, 81–150 (1987).
5. Schomaker, E. & Challa, G. Complexation of stereoregular poly(methyl methacrylates). 14. The basic structure of the stereocomplex of isotactic and syndiotactic poly(methyl methacrylate). *Macromolecules* **22**, 3337–3341 (1989).
6. Buter, R., Tan, Y. Y. & Challa, G. Radical polymerization of methyl methacrylate in the presence of stereoregular poly(methyl methacrylate). III. Influence of temperature. *J. Polym. Sci. Polym. Chem. Edn* **11**, 1013–1024 (1973).
7. Decher, G. in *Multilayer Thin Films* (eds Decher, G. & Schlenoff, J. B.) 1–46 (Wiley-VCH, Weinheim, 2003).
8. Serizawa, T. *et al.* Stepwise stereocomplex assembly of stereoregular poly(methyl methacrylate)s on a substrate. *J. Am. Chem. Soc.* **122**, 1891–1899 (2000).
9. Serizawa, T. *et al.* Stepwise assembly of isotactic poly(methyl methacrylate) and syndiotactic poly(methacrylic acid) on a substrate. *Langmuir* **16**, 7112–7115 (2000).
10. Hamada, K.-I. *et al.* Stepwise stereocomplex assembly of isotactic poly(methyl methacrylate) and syndiotactic poly(alkyl methacrylate)s on surfaces. *Langmuir* **17**, 5513–5519 (2001).
11. Serizawa, T., Hamada, K.-I., Kitayama, T. & Akashi, M. Recognition of stereoregular polymers using structurally-regulated ultrathin polymer films. *Angew. Chem. Int. Edn Engl.* **42**, 1118–1121 (2003).
12. Lohmeyer, J. H. G. M., Kransen, G., Tan, Y. Y. & Challa, G. Stereocoordination between poly(methyl methacrylate) and poly(methacrylic acid). *Polym. Lett. Edn* **13**, 725–729 (1975).

13. Lohmeyer, J. H. G. M., Tan, Y. Y., Lako, P. & Challa, G. Stereoselective association between isotactic poly(methyl methacrylate) and syndiotactic poly(methacrylic acid). *Polymer* **19**, 1171–1175 (1978).
14. Caruso, F., Caruso, R. A. & Möhwald, H. Nanoengineering of inorganic and hybrid hollow spheres by colloidal templating. *Science* **282**, 1111–1114 (1998).
15. Lohmeyer, J. H. G. M., Tan, Y. Y. & Challa, G. Polymerization of methacrylic acid in the presence of isotactic poly(methyl methacrylate) as possible template. *J. Macromol. Sci. A* **14**, 945–957 (1980).

Acknowledgements We thank K. Hatada and T. Kitayama for technological support of living anionic polymerization.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.A. (akashi@chem.eng.osaka-u.ac.jp).

Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends

Qiang Fu¹, Celeste M. Johanson¹, Stephen G. Warren¹ & Dian J. Seidel²

¹Department of Atmospheric Sciences, University of Washington, Seattle, Washington 98195, USA

²NOAA Air Resources Laboratory, Silver Spring, Maryland 20910, USA

From 1979 to 2001, temperatures observed globally by the mid-tropospheric channel of the satellite-borne Microwave Sounding Unit (MSU channel 2), as well as the inferred temperatures in the lower troposphere, show only small warming trends of less than 0.1 K per decade (refs 1–3). Surface temperatures based on *in situ* observations however, exhibit a larger warming of ~ 0.17 K per decade (refs 4, 5), and global climate models forced by combined anthropogenic and natural factors project an increase in tropospheric temperatures that is somewhat larger than the surface temperature increase^{6–8}. Here we show that trends in MSU channel 2 temperatures are weak because the instrument partly records stratospheric temperatures whose large cooling trend⁹ offsets the contributions of tropospheric warming. We quantify the stratospheric contribution to MSU channel 2 temperatures using MSU channel 4, which records only stratospheric temperatures. The resulting trend of reconstructed tropospheric temperatures from satellite data is physically consistent with the observed surface temperature trend. For the tropics, the tropospheric warming is ~ 1.6 times the surface warming, as expected for a moist adiabatic lapse rate.

The inconsistency between the trends at the surface and in the troposphere, traceable to the pioneering work of ref. 10, has raised questions about the ability of current global climate models (GCMs) to predict climate changes, the reliability of the observational data used to derive temperature trends, and the reality of human-induced climate change^{4,11–15}. It is generally agreed that the warming trend in global-mean surface temperature observations during the past 20 years is real and at least partly of anthropogenic origin^{4,12}. This increase of temperature is supported by observations of a reduction of snow cover and sea ice, thawing of permafrost, changes in freeze/thaw dates of lake and river ice, ocean warming and sea-level rise, and other related environmental changes⁴. However, the situation is less clear for tropospheric temperatures. Balloon-borne radiosondes have been the principal tool for atmospheric profiling. From 1979 to 2001, the trends of tropospheric-layer temperature between 850 and 300 hPa, as derived from different radiosonde data sets³, range from -0.03 to $+0.04$ K per decade. The radiosondes have limited spatial coverage, particularly

over large parts of the oceans, and are subject to a host of complications, including changing instrument types and observation practices^{16,17}, which confound analyses of climate trends.

The MSU, since 1979, and its successor, the Advanced MSU (AMSU), from 1998, provide a global measure of temperature for several atmospheric layers from NOAA polar-orbiting satellites. Although the original purpose of MSU measurements was to improve weather forecasts, a continuing data-analysis effort has been made to satisfy climate research requirements of homogeneity and calibration^{1,2,10,18–24}. Several important non-climatic influences have been identified and removed, including diurnal temperature biases related to local sampling times of the satellite and their changes over its lifetime, errors in the MSU calibration, and biases due to decay of the satellite orbits. Recent analyses of MSU channel 2 (most sensitive to the mid-troposphere) by the University of Alabama at Huntsville (UAH) and the Remote Sensing Systems (RSS) teams find temperature trends of 0.01 K per decade¹ and 0.1 K per decade², respectively, during 1979–2001. This trend difference is mainly due to differences in data adjustments related to instrument calibration and diurnal drift correction². The purpose of this Letter is not to reconcile the trend differences between these two research teams, but to address the question of whether the MSU data really imply small or negligible tropospheric warming over the past two decades. We argue that the trends reported by both teams for the ‘mid-tropospheric’ channel are substantially smaller than the actual trend of the mid-tropospheric temperature.

To infer the temperature of the mid-troposphere, we use two microwave channels: MSU channels 2 and 4 (or AMSU channels 5 and 9). Their weighting functions are shown in Fig. 1a. The weighting function for MSU channel 2 (or AMSU channel 5) peaks at ~550 hPa (~4.5 km). Thus the MSU Channel 2 brightness temperatures (T_2) have often been used to represent mid-tropospheric temperatures^{1,2,3,8,15,22,24}. The MSU channel 4 (or AMSU channel 9), whose weighting function peaks at ~85 hPa (~18 km), has been used to represent stratospheric temperatures^{1,3,8,15}.

As ~85% of the signal for T_2 comes from the troposphere and surface, it is not a bad approximation to say that the seasonal and interannual variations of mean deep-layer temperature in the troposphere can be well represented by T_2 . However, this might not be the case for trends. Figure 1b shows simulated changes of T_2 owing to changes of tropospheric and stratospheric temperatures, and indicates that T_2 remains constant when the changes in tropospheric and stratospheric temperatures have a ratio of about -1/5. This is because the vertical integral of the weighting function from the surface to tropopause, near 200 hPa, is about 5 times the integral above the tropopause. For example, if the tropospheric temperature trend were 0.15 K per decade and the stratospheric trend were -0.8 K per decade, the trend of T_2 would be close to zero. Intriguingly, this scenario resembles what may actually be the case in the atmosphere. The temperature trend in the lower stratosphere (15 to 23 km), as obtained from radiosondes and satellite observations, is about -0.5 to -0.9 K per decade for the last 20 years^{1,3,4,9}. Therefore, because of the combined influence of stratospheric and tropospheric changes, T_2 trends are not an ideal indicator of global climate change. To derive the tropospheric temperature trends, the effects of stratospheric cooling on T_2 must be taken into account.

Although a stratospheric influence on the T_2 trend has long been recognized^{20,25,26}, it has never been well quantified. The UAH team created a synthetic channel called T_{2LT} , where LT means ‘low-middle troposphere’^{1,19,25}, by subtracting signals at different view-angles of MSU channel 2. However, this approach amplifies noises, increases satellite inter-calibration biases, and may introduce other complications involving effects of changes in surface emissivity and of mountainous terrain^{2,19,20,24–28}. For these reasons, the T_{2LT} record is now receiving less attention than the better-calibrated T_2 record^{2,15,22,24}. Here we develop an alternative method to remove the stratospheric contribution, which should be free of the compli-

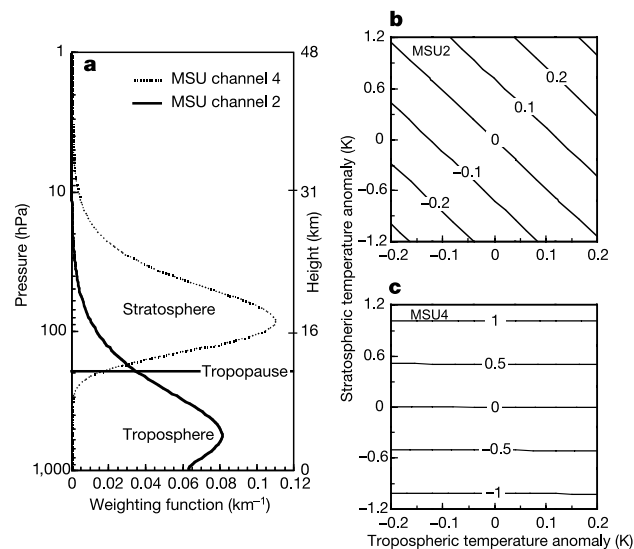


Figure 1 Atmospheric weighting functions and brightness temperature responses. **a**, Weighting function profiles for MSU channels 2 and 4 over ocean^{1,2}. The boundary between the troposphere and the stratosphere (the tropopause) is shown at 200 hPa. The satellite-observed brightness temperature, T_b , can be expressed in the form:

$$T_b = T_s W_s + \int_0^\infty T(z) W(z) dz$$

where T_s is the surface temperature, W_s the surface contribution factor, $T(z)$ the atmospheric temperature profile, and $W(z)$ the weighting function. Thus the weighting function describes the relative contributions of atmospheric temperatures at different heights to the brightness temperatures observed by the satellite. **b**, Responses of MSU channel 2 brightness temperature to changes in stratospheric and tropospheric temperatures assuming a US Standard Atmosphere and a surface emissivity of 0.5. Contours and axes are both labelled in K. **c**, Same as **b** but for MSU channel 4.

cations afflicting T_{2LT} , by making use of data from MSU channel 4.

The MSU channel 4 brightness temperature (T_4) is sensitive mainly to stratospheric temperature changes (Fig. 1c), so it can be used to remove the contribution of the stratosphere to T_2 . We define the free-tropospheric temperature as the mean temperature between 850 and 300 hPa ($T_{850-300}$; ref. 3). We derive this temperature from the measured brightness temperatures of MSU channels 2 and 4, as:

$$T_{850-300} = a_0 + a_2 T_2 + a_4 T_4 \quad (1)$$

To obtain these three coefficients, we use global-, hemispheric- and tropical-average monthly temperature anomaly profiles from radiosonde observations at 87 stations, for the period 1958–97 (ref. 17). The radiosonde data at the surface and at 15 pressure levels between 1,000 and 10 hPa are used to derive temperature anomalies for the 850–300-hPa layer as well as for MSU channels 2 and 4 (ref. 3). The coefficients in equation (1) are then obtained by least-squares regression (see Supplementary Table 1 for the values). For global-average anomalies, a_2 is 1.156 and a_4 is -0.153. The effective vertical weighting function for $T_{850-300}$ (that is, $a_2 W_2 + a_4 W_4$, where $W_{2,4}$ are the physical weighting functions for $T_{2,4}$) peaks at the same level as T_2 but is 15% larger. In the stratosphere it is negative above ~100 hPa and positive below, so that the integrated contribution of the stratosphere becomes near-zero. The effective weighting function may have a negative part²⁵; this is different from the physical weighting function, which must be positive everywhere.

The success of equation (1) in predicting $T_{850-300}$ from T_2 and T_4 is shown in Supplementary Fig. 1. The global-average anomalies of 850–300-hPa layer temperature, as derived from the radiosonde-simulated T_2 and T_4 , closely follow those directly observed by radiosondes for the period 1958–79. The correlation coefficient is 0.984, with a root-mean-square error of 0.065 K. The trend differences are only about 0.001 K per decade. The T_4 time series can

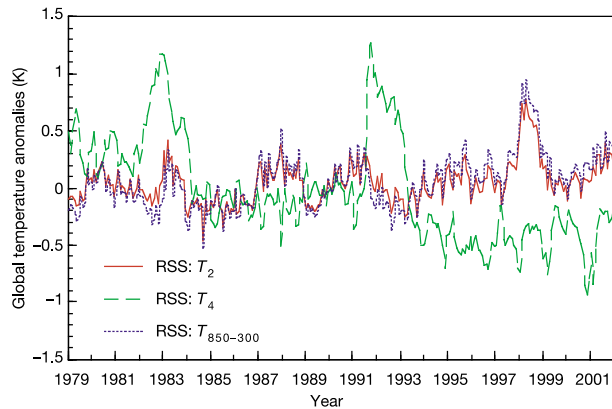


Figure 2 Time series of monthly mean, global-average temperature anomalies. Shown are the MSU channel 2 brightness temperature (T_2), MSU channel 4 brightness temperature (T_4) and MSU-derived 850–300-hPa layer temperature ($T_{850-300}$), based on the MSU data as analysed by the RSS team. All data are expressed as anomalies relative to climatological monthly means over 1985–94.

therefore be used to remove nearly all of the contribution of the stratosphere to T_2 in the trend analyses. This is because temperature variations in the stratosphere are vertically coherent and well correlated with T_4 .

We now apply equation (1) to satellite-observed time series of T_2 and T_4 from 1979 to 2001, as reported by UAH¹ and RSS², to derive $T_{850-300}$. The global anomaly time series of $T_{850-300}$ using RSS data is shown in Fig. 2, along with those of T_2 and T_4 . It is evident that the $T_{850-300}$ trend is more positive than the T_2 trend. (Similar results are obtained using UAH data, not shown here.)

Figure 3 shows the trends for T_2 (Fig. 3a) and MSU-derived $T_{850-300}$ (Fig. 3b) for the globe, Northern Hemisphere, Southern Hemisphere, and tropics (30°N – 30°S) using both the UAH and RSS datasets^{4,5}, as well as surface temperature trends based on *in situ* observations^{4,5}. The trends of 0.01 K per decade (UAH) and 0.1 K per decade (RSS) for global mean T_2 are substantially smaller than the surface temperature trend of 0.17 K per decade. In the Southern Hemisphere, the T_2 trend from UAH is actually negative. However, as shown in Fig. 3b, the global trends of $T_{850-300}$ are 0.09 K per decade (UAH) and 0.18 K per decade (RSS), which are about 0.08 K per decade larger than the corresponding T_2 trends. The trend difference between $T_{850-300}$ and T_2 for the tropics is smaller (~ 0.05 K per decade) because there the tropopause is higher and the stratospheric cooling is smaller, so the stratospheric influence is smaller.

Our analysis of the RSS data set suggests that over the past 22 years the global free troposphere has warmed at close to the same rate as the surface. The ratio of free-tropospheric temperature trend to surface temperature trend is ~ 1.1 for the globe and 1.6 for the tropics. In the tropics where the temperature follows the moist adiabatic lapse rate²⁹, this ratio should be larger than unity^{14,28,30}. GCM studies have predicted a global ratio of ~ 1.2 (ref. 8) and a tropical ratio of ~ 1.54 (ref. 14). Note that the RSS T_2 trend is also statistically consistent with a GCM prediction of T_2 (ref. 15).

Applying the stratospheric corrections to the UAH data set also enhances the mid-tropospheric temperature trends, but they are still smaller than the surface warming rate, particularly over the Southern Hemisphere (Fig. 3b). In addition, the UAH-reported T_{2LT} (bulk temperatures for the low–middle troposphere) seems to be inconsistent with the $T_{850-300}$ obtained from the UAH data. For example, over the period from 1979 to 2001, the UAH-reported T_{2LT} trend in the tropics is -0.01 K per decade, which is substantially smaller than the $T_{850-300}$ trend of 0.08 K per decade. (The trend of the difference time series between $T_{850-300}$ and T_{2LT} (that is, $T_{850-300} - T_{2LT}$), which is 0.09 ± 0.05 K per decade (the 95%

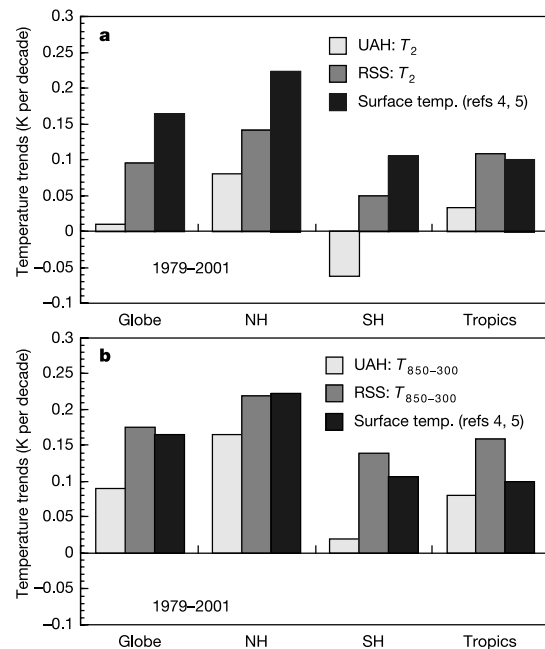


Figure 3 Trends in monthly mean temperature anomalies. **a**, MSU channel 2 brightness temperature trend and **b**, MSU-derived 850–300-hPa layer temperature trend for the globe, Northern Hemisphere (NH), Southern Hemisphere (SH) and tropics (30°N – 30°S). The results using both UAH (version 5) and RSS data sets are shown. See Supplementary Table 2 for these trend values and their error estimates. The surface temperature trends are also shown for comparison. The surface temperature trends for the globe, Northern Hemisphere and Southern Hemisphere are from ref. 5; the surface trend for the tropical region is from ref. 4.

confidence interval), is significantly different from zero at less than 0.1% significance level.) Also note that in the tropics the UAH T_{2LT} is cooling at -0.04 K per decade relative to the UAH T_2 . This apparent inconsistency may be attributed to complications involving the T_{2LT} retrieval, as well as to the techniques used by UAH to analyse the MSU channel 2 data, which could also explain the lack of agreement between GCM simulations⁸ and UAH results for the trend differences of either $T_s - T_{2LT}$, where T_s is the surface temperature, or $T_s - T_2$, despite their agreement for T_2 trends⁸.

In an independent analysis of MSU data, Vinnikov and Grody²⁴ found a large positive global trend of 0.22–0.26 K per decade for T_2 , which they used to represent the tropospheric temperature. But irrespective of the techniques used to analyse the data, T_2 is subject to the effects of stratospheric cooling. Assuming a stratospheric temperature trend of -0.5 K per decade^{1,3,4}, the Vinnikov–Grody T_2 trend translates to a $T_{850-300}$ trend of ~ 0.33 – 0.37 K per decade. This value is about twice as large as the surface warming globally. It also suggests a ratio of ~ 3 between tropospheric and surface temperature trends for the tropical region. These ratios seem large, and suggest that the technique Vinnikov and Grody used to analyse the satellite data may require further scrutiny. □

Received 8 December 2003; accepted 29 March 2004; doi:10.1038/nature02524.

- Christy, J. R. *et al.* Error estimates of version 5.0 of MSU-AMSU bulk atmospheric temperatures. *J. Atmos. Ocean. Technol.* **20**, 613–629 (2003).
- Mears, C. A., Schabel, M. C. & Wentz, F. J. A reanalysis of the MSU channel 2 tropospheric temperature record. *J. Clim.* **16**, 3650–3664 (2003).
- Seidel, D. J. *et al.* Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets. *J. Clim.* (in the press).
- Houghton, J. T. *et al.* in *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, London, 2001).
- Jones, P. D. & Moberg, A. Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Clim.* **16**, 206–223 (2003).
- Bengtsson, L., Roeckner, E. & Stendel, M. Why is the global warming proceeding much slower than expected? *J. Geophys. Res.* **104**, 3865–3876 (1999).

7. Santer, B. D. *et al.* Interpreting differential temperature trends at the surface and in the lower troposphere. *Science* **287**, 1227–1232 (2000).
8. Hansen, J. *et al.* Climate forcings in Goddard Institute for Space Studies SI2000 simulations. *J. Geophys. Res.* **107**, doi:10.1029/2001JD001143 (2002).
9. Ramaswamy, V. *et al.* Stratospheric temperature trends: Observations and model simulations. *Rev. Geophys.* **39**, 71–122 (2001).
10. Spencer, R. W. & Christy, J. R. Precise monitoring of global temperature trends from satellites. *Science* **247**, 1558–1662 (1990).
11. Hansen, J. *et al.* Satellite and surface temperature data at odds? *Clim. Change* **30**, 103–117 (1995).
12. Wallace, J. M. *et al.* *Reconciling Observations of Global Temperature Change* (National Academy Press, Washington DC, 2000).
13. Singer, S. F. Difficulty in reconciling global warming data. *Nature* **409**, 281 (2001).
14. Hegerl, G. C. & Wallace, J. M. Influence of patterns of climate variability on the difference between satellite and surface temperature trends. *J. Clim.* **15**, 2412–2428 (2002).
15. Santer, B. D. *et al.* Influence of satellite data uncertainties on the detection of externally forced climate change. *Science* **300**, 1280–1284 (2003).
16. Gaffen, D. J., Sargent, M. A., Habermann, R. E. & Lanzante, J. R. Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality. *J. Clim.* **13**, 1776–1796 (2000).
17. Lanzante, J. R., Klein, S. A. & Seidel, D. J. Temporal homogenization of monthly radiosonde temperature data. *J. Clim.* **16**, 224–262 (2003).
18. Christy, J. R., Spencer, R. W. & McNider, R. T. Reducing noise in the MSU daily lower-tropospheric global temperature dataset. *J. Clim.* **8**, 888–896 (1995).
19. Christy, J. R., Spencer, R. W. & Lobl, E. S. Analysis of the merging procedure for the MSU daily temperature time series. *J. Clim.* **11**, 2016–2041 (1998).
20. Wentz, F. J. & Schabel, M. Effects of orbital decay on satellite-derived lower-tropospheric temperature trends. *Nature* **394**, 661–664 (1998).
21. Christy, J. R., Spencer, R. W. & Braswell, W. D. MSU tropospheric temperature: Dataset construction and radiosonde comparisons. *J. Atmos. Ocean. Technol.* **17**, 1153–1170 (2000).
22. Prabhakara, C., Iacovazzi, J. R., Yoo, J. M. & Dalu, G. Global warming: Evidence from satellite observations. *Geophys. Res. Lett.* **27**, 3517–3520 (2000).
23. Mo, T., Goldberg, M. D. & Crosby, D. S. Recalibration of the NOAA microwave sounding unit. *J. Geophys. Res.* **106**, 10145–10150 (2001).
24. Vinnikov, K. Y. & Grody, N. C. Global warming trend of mean tropospheric temperature observed by satellites. *Science* **302**, 269–272 (2003).
25. Spencer, R. W. & Christy, J. R. Precision and radiosonde validation of satellite gridpoint temperature anomalies. Part II: A tropospheric retrieval and trends during 1979–1990. *J. Clim.* **5**, 858–866 (1992).
26. Hurrell, J. W. & Trenberth, K. E. Spurious trends in satellite MSU temperature from merging different satellite records. *Nature* **386**, 164–167 (1997).
27. Trenberth, K. E. & Hurrell, J. W. Reply to “How accurate are satellite ‘thermometers?’”. *Nature* **389**, 342–343 (1997).
28. Hurrell, J. W. & Trenberth, K. E. Difficulties in obtaining reliable temperature trends: Reconciling the surface and satellite Microwave Sounding Unit records. *J. Clim.* **11**, 945–967 (1998).
29. Stone, P. H. & Carlson, J. H. Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *J. Atmos. Sci.* **36**, 415–423 (1979).
30. Wentz, F. J. & Schabel, M. Precise climate monitoring using complementary satellite data sets. *Nature* **403**, 414–416 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. M. Wallace for discussions. We also thank J. M. Wallace, D. L. Hartmann, J. R. Holton, J. K. Angell and M. Free for their comments on the manuscript. This study was supported by the US DOE, NSF and NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Q.F. (qfu@atmos.washington.edu).

Partitioning of oxygen during core formation on the Earth and Mars

David C. Rubie, Christine K. Gessmann* & Daniel J. Frost

Bayerisches Geoinstitut, University of Bayreuth, D-95444 Bayreuth, Germany

* Present address: Gleiwitzerstrasse 3, D-55131 Mainz, Germany

Core formation on the Earth and Mars involved the physical separation of metal and silicate, most probably in deep magma oceans^{1–4}. Although core-formation models explain many aspects of mantle geochemistry, they have not accounted for the large differences observed between the compositions of the mantles of the Earth (~8 wt% FeO) and Mars (~18 wt% FeO) or the smaller

mass fraction of the martian core^{5–7}. Here we explain these differences as a consequence of the solubility of oxygen in liquid iron-alloy increasing with increasing temperature. We assume that the Earth and Mars both accreted from oxidized chondritic material. In a terrestrial magma ocean, 1,200–2,000 km deep, high temperatures resulted in the extraction of FeO from the silicate magma ocean owing to high solubility of oxygen in the metal. Lower temperatures of a martian magma ocean resulted in little or no extraction of FeO from the mantle, which thus remains FeO-rich. The FeO extracted from the Earth's magma ocean may have contributed to chemical heterogeneities in the lowermost mantle⁸, a FeO-rich D'' layer⁹ and the light element budget of the core^{10,11}.

The most significant differentiation event in the history of the Earth and other terrestrial planets was the separation of metal and silicate to form a metallic Fe-rich core and a silicate mantle. The geochemical consequences of this process have been studied extensively in recent years to explain the geochemistry of the Earth's mantle and the physical mechanisms of core formation and accretion. Extensive melting of the Earth and the formation of a deep magma ocean as a consequence of one or more giant impacts probably facilitated metal–silicate separation^{12,13}. The content of moderately siderophile (metal-loving) elements, such as Ni and Co, in the Earth's mantle can be explained by the separation of liquid metal and silicate in a magma ocean at least 700 km deep^{1–3}. This conclusion is based on experimental studies of the partitioning of siderophile elements between liquid metal and liquid silicate at high pressure. Here we consider the partitioning of oxygen between metal and silicate in a magma ocean with the aim of understanding how the FeO content and oxidation state of planetary mantles are affected by core formation.

We studied the solubility of oxygen in liquid Fe–Ni alloy at 9 and 18 GPa, 2,173–2,673 K and oxygen fugacities (f_{O_2}) 1.1 to 3.6 log units below the iron–wüstite buffer by equilibrating the samples with magnesiowüstite in a multianvil apparatus. Details of the starting materials and experimental and analytical methods have been described previously¹⁴. Over the range of experimental conditions, the solubility varies from below detection limit to 1.28 wt% (Fig. 1a, Table 1). The results show that oxygen solubility increases with increasing temperature and oxygen fugacity (Fig. 1a). To determine the effects of pressure and temperature independently of f_{O_2} , we calculate the distribution coefficient, K_d , for the partitioning of oxygen between liquid Fe-alloy and magnesiowüstite:

$$K_d = \frac{X_{\text{O}}^{\text{met}} X_{\text{Fe}}^{\text{met}}}{X_{\text{FeO}}^{\text{mw}}} \quad (1)$$

where $X_{\text{O}}^{\text{met}}$, $X_{\text{Fe}}^{\text{met}}$ and $X_{\text{FeO}}^{\text{mw}}$ are the mole fractions of oxygen in metal, Fe in metal and FeO in magnesiowüstite respectively. As shown in Fig. 1b, the distribution coefficient, and therefore the oxygen solubility at constant f_{O_2} , decreases with increasing pressure. These trends are consistent with previous results obtained at 5–25 GPa and 2,073–2,773 K but over a relatively restricted f_{O_2} range¹⁵. Note that the results are contrary to early predictions that the solubility of oxygen in liquid Fe should increase with increasing pressure¹⁶.

To extrapolate the oxygen solubility data as a function of pressure (P) and temperature (T), we use:

$$RT \ln K_d = -\Delta H + T\Delta S - P\Delta V \quad (2)$$

where ΔH , ΔS and ΔV are the changes in enthalpy, entropy and volume, respectively, for the oxygen exchange reaction and R is the gas constant. We fitted equation (2) to the data, including five experimental results obtained up to 25 GPa by ref. 15. Results of the fit, shown in Fig. 1a and b, give $\Delta H = 153,000 (\pm 29,000) \text{ J mol}^{-1}$, $\Delta S = 50.9 (\pm 12.4) \text{ J K}^{-1} \text{ mol}^{-1}$ and $\Delta V = 1,448 (\pm 440) \text{ J GPa}^{-1}$ ($1.448 \pm 0.44 \text{ cm}^3 \text{ mol}^{-1}$).

Extrapolations of the solubility results to higher pressures and

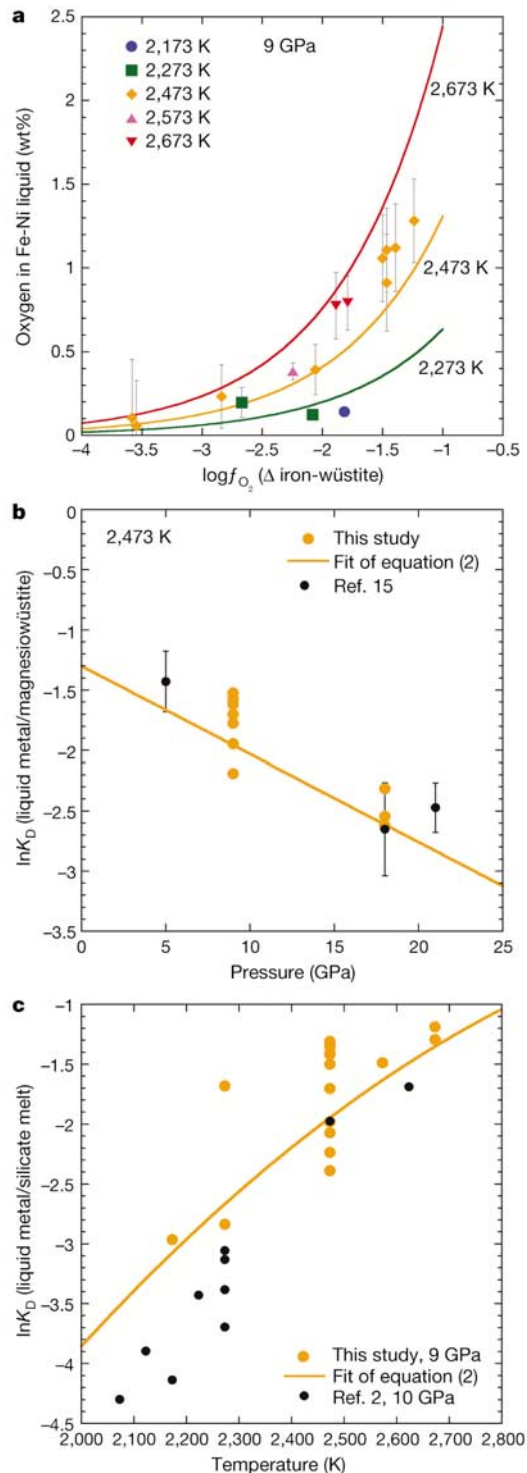


Figure 1 Oxygen partitioning results. **a**, **b**, Experimental results plotted as oxygen content in liquid Fe-alloy as a function of oxygen fugacity (f_{O_2}), relative to the iron-wüstite buffer, at 9 GPa and 2,173–2,673 K (**a**) and $\ln K_d$ as a function of pressure at 2,473 K; see equation (1) (**b**). At constant f_{O_2} , oxygen solubility increases with increasing temperature and decreases with increasing pressure. **c**, $\ln K_d$ for the partitioning of oxygen between liquid metal alloy and silicate melt calculated from liquid metal alloy/magnesiowüstite partitioning data, using equation (3). Values of $\ln K_d$ for oxygen partitioning between liquid Fe-Ni-S alloy and silicate melt, calculated from the results of ref. 2, are shown for comparison. The line is the fit of equation (2) to our data, modified using equation (3).

temperatures, using equation (2), are shown in Fig. 2. At moderate pressures (5–10 GPa) oxygen solubility is predicted to reach 5–15 wt% at 3,500–4,000 K. However, as pressure exceeds ~ 40 GPa, oxygen solubility becomes low, even at very high temperatures.

We examine the consequences of the partitioning of oxygen between liquid Fe-Ni metal and silicate liquid in a magma ocean. To determine the partitioning of oxygen between liquid metal and peridotite liquid in a magma ocean, we use the empirical relationship:

$$X_{FeO}^{mw} = 1.148X_{FeO}^{sil} + 1.319(X_{FeO}^{sil})^2 \quad (3)$$

where X_{FeO}^{sil} is the mole fraction of FeO in silicate liquid. Equation (3) has been derived using the results of melting studies of peridotite and Fe-rich peridotite bulk compositions^{17,18}. Figure 1c shows the distribution coefficient K_d for the partitioning of oxygen between liquid Fe-alloy and silicate melt calculated from our liquid Fe-alloy/magnesiowüstite partitioning data using equation (3). These data and the recalculated fit of equation (2) are compared with previously published results on the partitioning of oxygen between liquid Fe-Ni-S alloy and silicate melt^{2,19}. The similar temperature dependence and absolute values of K_d observed indicates that the addition of up to 30 wt% sulphur has a very minor effect on oxygen solubility (Fig. 1c).

For a given bulk composition, the equilibrium compositions and mass fractions of coexisting metal and silicate are calculated by mass balance using equations (2) and (3). We consider the geochemical consequences of metal-silicate separation using a polybaric metal-silicate fractionation model²⁰. We assume that the temperature at the base of the magma ocean is defined by the peridotite liquidus^{21,22} and determine temperature as a function of depth by calculating magma ocean adiabats²⁰.

We assume that the material that accreted to form Earth and Mars contained the major elements Si, Al, Mg, Fe and Ca in chondritic (C1) proportions before metal-silicate separation²³. The oxygen contents of chondritic meteorites are variable and poorly defined²⁴ but, with the exception of enstatite chondrites, the silicate component is generally considerably more oxidized than the Earth's mantle. We therefore vary the amount of oxygen in the starting composition such that between 50 and 80 wt% of the total Fe is present initially as reduced metal, with the remainder being present as an FeO component. These two limits correspond to the non-metallic (silicate) fraction of the chondritic material containing 22 and 10 wt% FeO respectively (for comparison, the Earth's mantle contains ~ 8 wt% FeO).

The FeO content of the residual silicate mantle, after the metal has segregated, is calculated as a function of magma ocean depth for

Table 1 Experimental conditions and results

Run number	Temperature (K)	Pressure (GPa)	$\log(f_{O_2})$	Oxygen (wt%)	$\ln(K_d)$
1,549	2,173	9	-1.82 (0.01)	0.14 (0.01)	-3.218
1,558	2,273	9	-2.08 (0.01)	0.12 (0.01)	-3.077
1,358	2,273	9	-2.67 (0.09)	0.20 (0.09)	-1.903
1,203	2,473	9	-3.59 (0.35)	0.11 (0.35)	-1.525
1,186	2,473	9	-2.06 (0.15)	0.39 (0.15)	-1.945
1,316	2,473	9	-3.55 (0.27)	0.06 (0.27)	-2.189
1,210	2,473	9	-1.47 (0.29)	0.91 (0.29)	-1.776
1,218	2,473	9	-2.84 (0.19)	0.23 (0.19)	-1.574
1,383	2,473	9	-1.47 (0.25)	1.11 (0.25)	-1.585
1,194	2,473	9	-1.24 (0.25)	1.28 (0.25)	-1.704
1,371	2,473	9	-1.39 (0.26)	1.12 (0.26)	-1.7
1,352	2,473	9	-1.50 (0.26)	1.06 (0.26)	-1.62
1,547	2,573	9	-2.25 (0.05)	0.38 (0.05)	-1.721
1,312	2,673	9	-1.89 (0.20)	0.78 (0.20)	-1.435
1,551	2,673	9	-1.79 (0.16)	0.79 (0.16)	-1.549
1,273	2,473	18	-2.32 (0.20)	0.14 (0.20)	-2.621
1,275	2,473	18	-2.08 (0.18)	0.27 (0.18)	-2.314
1,285	2,473	18	-1.12 (0.25)	0.62 (0.25)	-2.545

A full list of sample analyses is provided in the Supplementary Information. $\log(f_{O_2})$ is calculated relative to the iron-wüstite buffer; K_d is defined in equation (1).

Earth and Mars (Fig. 3). In the case of Earth, the temperatures of magma oceans with depths <800 km are too low for significant amounts of oxygen to dissolve in the metal phase. Consequently, the initial FeO content of the silicate fraction of the chondritic material is changed only slightly by metal segregation (Fig. 3a). As the magma ocean depth is increased beyond 800 km, however, the amount of oxygen that partitions into the metal increases (owing to higher temperatures) with the consequence that the FeO content of the residual silicate mantle becomes increasingly reduced. Using the simple extrapolation model of equation (2), a magma ocean with a depth of ~1,800 km gives the current FeO content (8 wt%) and Fe/(Fe + Mg) ratio (0.1) of the Earth's mantle, assuming that the starting chondritic material was relatively oxidized with an initial FeO content of the order of 18–22 wt% (Fig. 3a). Considering the model uncertainties, the magma ocean depth could be as low as 1,200 km. A magma ocean with a depth of $\geq 1,200$ km has also been proposed on the basis of siderophile element partitioning studies^{2,25–27}.

In the case of Mars, the model predicts little change in the FeO content of the residual mantle as a result of core formation, irrespective of magma ocean depth (Fig. 3b). This is because the pressures and temperatures of a martian magma ocean are much lower than on Earth⁴; consequently, little or no oxygen dissolves in the liquid metal because of the low temperatures. Our results therefore show that Earth and Mars could have accreted from material with the same initial FeO content (~18 wt%) and that

the different FeO contents of their mantles are a consequence of the partitioning behaviour of oxygen during core formation. The model also explains the different mass fractions of the cores of Earth and Mars because more metal segregates from the magma ocean on Earth (~28 wt%) than on Mars (~18 wt%).

Although the model requires large extrapolations of the experimental data, the associated uncertainties do not affect our conclusions for Earth and Mars (see Supplementary Information). Using existing thermodynamic data^{28,29} on pure Fe liquid, oxygen and iron oxide we can calculate a value for ΔS that is in good agreement with our determined value. This calculation indicates that the inclusion of heat capacity functions to determine the temperature dependence of ΔS even up to 6,000 K has a very minor effect on the results of the model. The extrapolation assumes a constant positive value for ΔV of the metal–oxide equilibrium, which means that oxygen solubility decreases with pressure. This trend could reverse, however, at pressures higher than those of our experiments if the value for ΔV becomes negative as a result of compressibility or phase transformation³⁰. The martian mantle, however, extends only to pressures of 25 GPa and is therefore within

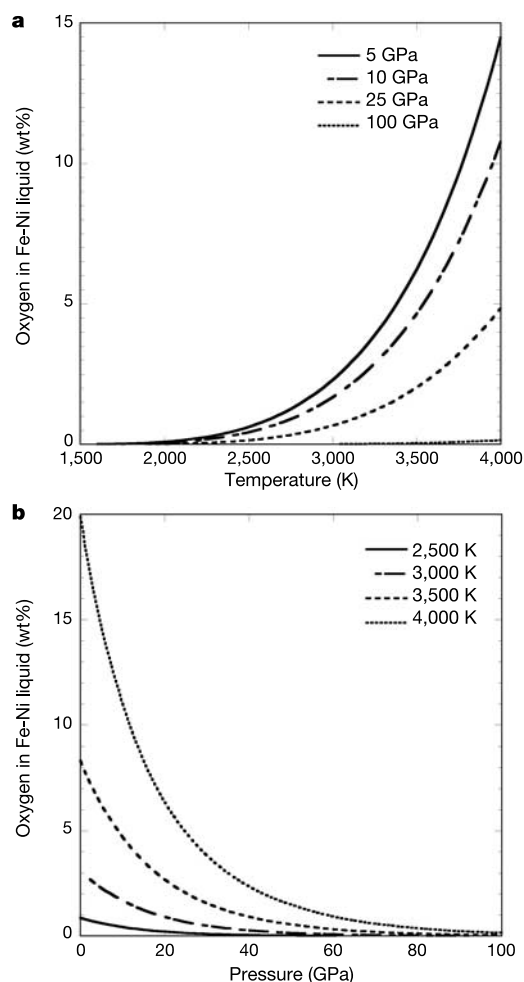


Figure 2 Extrapolations of the oxygen solubility data to higher temperatures and pressures (**a** and **b**) using equation (2). The oxygen fugacity is fixed at two log units below the iron–wüstite buffer.

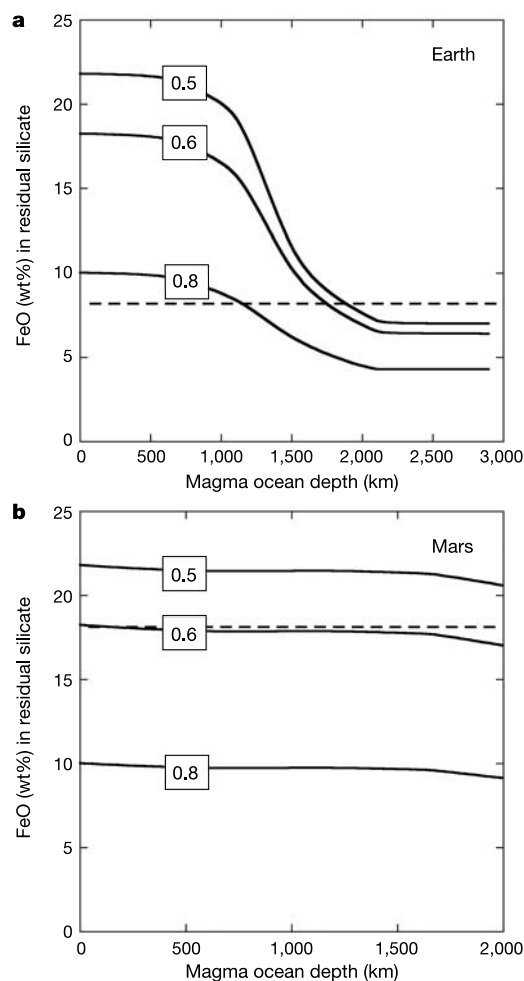


Figure 3 Results of the metal–silicate separation model for Earth (**a**) and Mars (**b**). The FeO contents of the residual silicate mantles (after metal segregation) are shown as a function of magma ocean depth. In each case, the three curves correspond to different initial oxygen contents and are labelled with the weight fraction of Fe that is initially present as metal; the three values correspond to 22, 18 and 10 wt% FeO in the silicate fraction, respectively. The results show that metal–silicate separation in a deep magma ocean would strongly reduce the FeO content of the silicate mantle on Earth, whereas on Mars metal–silicate separation has little effect. The horizontal broken lines indicate the FeO contents of the mantles of Earth and Mars.

the experimental pressure range. If oxygen solubility does increase at pressures higher than 25 GPa, this would reduce the magma ocean depth required on the Earth to satisfy the FeO content of the present-day mantle and make it more likely that oxygen remained dissolved in the metal as it formed the core.

An important consequence for the Earth is that the metal that segregated at the base of the magma ocean would have contained a significant amount of oxygen (for example, 7–8 wt%). As a result of increasing pressure, the solubility of oxygen in Fe-alloy is predicted to decrease with depth below the magma ocean (Fig. 2). There are three possibilities for the fate of the dissolved oxygen, all of which could be important. (1) FeO precipitated from liquid metal between the base of the magma ocean and the proto-core, thus creating chemical heterogeneities that may still exist in the lowermost mantle⁸. (2) Oxygen was transported by liquid metal to the core, where FeO precipitated to form the D'' layer at the core–mantle boundary⁹. (3) At least part of the oxygen may still be dissolved in the core (up to 8 wt% according to our model), thus contributing significantly to its light-element budget^{10,11}. If oxygen does become more soluble in liquid metal at pressures higher than our experimental conditions, then that would tend to favour this final possibility. In contrast, the martian core is unlikely to contain any dissolved oxygen, which is consistent with its probable high sulphur content⁷. □

Received 6 October 2003; accepted 8 March 2004; doi:10.1038/nature02473.

- Li, J. & Agee, C. B. Geochemistry of mantle–core differentiation at high pressure. *Nature* **381**, 686–689 (1996).
- Li, J. & Agee, C. B. The effect of pressure, temperature, oxygen fugacity and composition on partitioning of nickel and cobalt between liquid Fe–Ni–S alloy and liquid silicate: Implications for the Earth's core formation. *Geochim. Cosmochim. Acta* **65**, 1821–1832 (2001).
- Righter, K., Drake, M. J. & Yaxley, G. Prediction of siderophile element/silicate partition coefficients to 20 GPa and 2800 °C: the effects of pressure, temperature, oxygen fugacity, and silicate and metallic melt compositions. *Phys. Earth Planet. Inter.* **100**, 115–134 (1997).
- Righter, K., Hervig, R. L. & Kring, D. A. Accretion and core formation on Mars: molybdenum contents of melt inclusion glasses in three SNC meteorites. *Geochim. Cosmochim. Acta* **62**, 2167–2177 (1998).
- Anderson, D. L. *Theory of the Earth* (Blackwell, Boston, 1989).
- Dreibus, G. & Wanke, H. Mars, a volatile-rich planet. *Meteoritics* **20**, 367–381 (1985).
- McSween, H. Y. Jr What we have learnt about Mars from SNC meteorites. *Meteoritics* **29**, 757–779 (1994).
- Kellogg, L. H., Hager, B. H. & van der Hilst, R. D. Compositional stratification in the deep mantle. *Science* **283**, 1881–1884 (1999).
- Garnero, E. J., Revenaugh, J., Williams, Q., Lay, T. & Kellogg, L. H. in *The Core–Mantle Boundary Region* (eds Gurnis, M., Wyssession, M. E., Knittle, E. & Buffett, B. A.) 319–334 (Geodynamics Series 28, Am. Geophys. Union, Washington, 1998).
- Poirier, J.-P. Light elements in the Earth's outer core: A critical review. *Phys. Earth Planet. Inter.* **85**, 319–337 (1994).
- Hillgren, V., Gessmann, C. K. & Li, J. in *Origin of the Earth and Moon* (eds Canup, R. & Righter, K.) 245–263 (Univ. Arizona Press, Tucson, 2000).
- Melosh, H. J. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 69–83 (Oxford Univ. Press, Oxford, 1990).
- Stevenson, D. J. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 231–249 (Oxford Univ. Press, Oxford, 1990).
- Gessmann, C. K. & Rubie, D. C. The effect of temperature on the partitioning of Ni, Co, Mn, Cr and V at 9 GPa and constraints on formation of the Earth's core. *Geochim. Cosmochim. Acta* **62**, 867–882 (1998).
- O'Neill, H. S. C., Canil, D. & Rubie, D. C. Metal-oxide equilibria to 2500 °C and 25 GPa: implications for core formation and the light component in the Earth's core. *J. Geophys. Res.* **103**, 12239–12260 (1998).
- Kato, T. & Ringwood, A. E. Melting relationships in the system Fe–FeO at high pressures: implications for the composition and formation of the Earth's core. *Phys. Chem. Miner.* **16**, 524–538 (1989).
- Trønnes, R. G. Melting relations and major element partitioning in an oxidized bulk Earth model composition at 15–26 GPa. *Lithos* **53**, 233–245 (2000).
- Trønnes, R. G. & Frost, D. J. Peridotite melting and mineral–melt partitioning of major and minor elements at 22–24.5 GPa. *Earth Planet. Sci. Lett.* **197**, 117–131 (2002).
- Li, J. & Agee, C. B. Element partitioning constraints on the light element composition of the Earth's core. *Geophys. Res. Lett.* **28**, 81–84 (2001).
- Rubie, D. C., Melosh, H. J., Reid, J. E., Liebske, C. & Righter, K. Mechanisms of metal–silicate equilibration in the terrestrial magma ocean. *Earth Planet. Sci. Lett.* **205**, 239–255 (2003).
- Herzberg, C. & Zhang, J. Melting experiments on anhydrous peridotite KLB-1: Compositions of magmas in the upper mantle and transition zone. *J. Geophys. Res.* **101**, 8271–8295 (1996).
- Ohtani, E. Melting temperature distribution and fractionation in the lower mantle. *Phys. Earth Planet. Inter.* **33**, 12–25 (1983).
- McDonough, W. F. & Sun, S.-s. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- O'Neill, H. S. C. & Palme, H. in *The Earth's Mantle* (ed. Jackson, I.) 3–126 (Cambridge Univ. Press, Cambridge, 1998).
- Chabot, N. L. & Agee, C. B. Core formation in the Earth and Moon: new experimental constraints from V, Cr and Mn. *Geochim. Cosmochim. Acta* **67**, 2077–2091 (2003).

- Gessmann, C. K. & Rubie, D. C. The origin of the depletions of V, Cr, and Mn in the mantles of the Earth and Moon. *Earth Planet. Sci. Inter.* **184**, 95–107 (2000).
- Bouhifd, M. A. & Jephcoat, A. P. The effect of pressure on partitioning of Ni and Co between silicate and iron-rich metal liquids: a diamond–anvil cell study. *Earth Planet. Sci. Inter.* **209**, 245–255 (2003).
- Chase, M. W. J. et al. JANAF thermochemical tables. *J. Phys. Chem. Ref. Data* **14** (suppl. 1) (1985).
- Sundman, B. An assessment of the Fe–O system. *J. Phase Equil.* **12**, 127–140 (1991).
- Fei, Y. & Mao, H.-K. In situ determination of the NiAs phase of FeO at high pressure and temperature. *Science* **266**, 1678–1680 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Fischer, G. Herrmannsdörfer, D. Krause and H. Schulze for technical assistance. The German Science Foundation (DFG) supported this research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.C.R. (dave.rubie@uni-bayreuth.de).

Convergent evolution in mechanical design of lamnid sharks and tunas

Jeanine M. Donley¹, Chugey A. Sepulveda¹, Peter Konstantinidis², Sven Gemballa² & Robert E. Shadwick¹

¹Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0202, USA

²Department of Zoology, University of Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany

The evolution of 'thunniform' body shapes in several different groups of vertebrates, including whales, ichthyosaurs¹ and several species of large pelagic fishes² supports the view that physical and hydromechanical demands provided important selection pressures to optimize body design for locomotion during vertebrate evolution. Recognition of morphological similarities between lamnid sharks (the most well known being the great white and the mako) and tunas has led to a general expectation that they also have converged in their functional design; however, no quantitative data exist on the mechanical performance of the locomotor system in lamnid sharks. Here we examine the swimming kinematics, *in vivo* muscle dynamics and functional morphology of the force-transmission system in a lamnid shark, and show that the evolutionary convergence in body shape and mechanical design between the distantly related lamnids and tunas is much more than skin deep; it extends to the depths of the myotendinous architecture and the mechanical basis for propulsive movements. We demonstrate that not only have lamnids and tunas converged to a much greater extent than previously known, but they have also developed morphological and functional adaptations in their locomotor systems that are unlike virtually all other fishes.

During their 400 million years of independent evolution, sharks and bony fishes have diverged in many fundamental aspects of their anatomy and physiology. However, two groups of dominant open-ocean predators, the lamnid sharks and the tunas, even when looked at superficially, display remarkably similar morphological specializations related to locomotion^{3–12} (Fig. 1a). The shared characteristics in these distantly related groups that distinguish them from virtually all other fish have arisen independently, probably as the result of selection for fast and continuous locomotion. Moreover, in both lamnids and tunas, the aerobic (red) musculature that powers cruise swimming is concentrated in a more medial (closer to the

backbone) and anterior position compared with the relatively uniform and superficial position in other fishes; the body temperature is elevated above that of the surrounding water, facilitated by counter-current heat exchangers associated with the internal red muscle; and the muscle segments (myotomes) are highly elongated³.

Recent studies on tunas revealed a host of unique functional adaptations in their locomotor system that distinguish tunas from

all other bony fishes^{4,13}. Similar investigations on swimming lamnid sharks are lacking because of the difficulty in handling such large and dangerous predators; thus, the dynamic properties of the lamnid locomotor system remain unknown. This paper presents *in vivo* quantitative measurements of swimming kinematics and muscle dynamics, and analysis of the morphology of the force-transmission system in a lamnid shark.

First, we examined the kinematics of steady swimming in the shortfin mako shark (*Isurus oxyrinchus*). Lateral displacement of the dorsal midline as a function of body position was calculated from dorsal video images of mako sharks swimming under controlled conditions in a swim tunnel (Fig. 1b; see also Supplementary Video). Descriptions of undulatory swimming modes in fishes are based on the proportion of the body that participates in the lateral thrust-producing movements¹⁴, and can be distinguished by differences in patterns of lateral displacement along the body, as shown in Fig. 1c. Compared with other teleosts, tunas exhibit the least undulatory (thunniform) mode of locomotion, in which lateral movements are largely confined to the caudal region where body mass is reduced by tapering. The lateral displacement data presented here show that the lamnid shark kinematically resembles tuna more than other sharks¹⁵ or subcarangiform teleosts: the degree of lateral motion along the mako shark's body from $\sim 0.4L$ to $\sim 0.8L$ (where L is total body length) is relatively small, demonstrating that lamnids swim using a thunniform-like mode. The amplitude of lateral motion increases substantially beyond $\sim 0.8L$ where the body tapers to the narrow caudal peduncle. Lamnids, like tunas, have the least lateral motion in the mid-body region where the bulk of the muscle resides and have a reduced body mass in the caudal region where lateral amplitudes are high, both being features that match predictions for enhancing hydromechanical efficiency of swimming⁹.

Because lamnids have both internal red muscle and a thunniform-like swimming mode, we tested the possibility that shortening of the red muscle would be physically uncoupled from deformation of the adjacent skin, backbone and white muscle, a unique functional property of red muscle that has been observed so far only in tunas. Whereas in most teleosts superficial red muscle fibres contract sequentially to cause a posteriorly travelling wave of local body bending^{16–18}, the internal position of red muscle in tunas allows them to abandon this pattern of undulation and adopt a novel mechanism to project red muscle action to posterior regions of the body^{4,19}, thereby facilitating thunniform kinematics. In the mako shark we used sonomicrometry to record instantaneous muscle segment length changes during steady swimming as well as during passive, simulated swimming movements induced under anaesthesia. The temporal relationship between red and white muscle shortening was measured to determine whether the action of these two muscle masses is synchronized, as in most fish, or uncoupled, as in tunas. During passive swimming movements, peaks in red muscle strain (that is, relative length change) were in phase with peaks in adjacent white muscle strain (Fig. 2a). Thus, when the body bends passively red muscle shortens synchronously with the surrounding white muscle and skin, as one would expect. In contrast, during active swimming the peaks in red muscle strain were delayed relative to peaks in white muscle strain (Fig. 2b). By cross-correlation analysis we determined that the mean phase shift between simultaneous recordings of red and white muscle strain was 90 ms (or $\sim 10\%$ of the tail-beat cycle), with one individual as high as 174 ms ($\sim 17\%$ of tail-beat cycle). The observed phase shift indicates that during steady swimming the red muscle is indeed physically uncoupled from the surrounding tissues and contracts in phase with body bending at a more posterior location.

On the basis of a wave velocity of about 11 s^{-1} in sharks swimming with a tail-beat frequency of about 1 Hz, the red muscle in the mid-body region will be shortening in phase with bending of the backbone 10–17% closer to the tail. A consequence is that

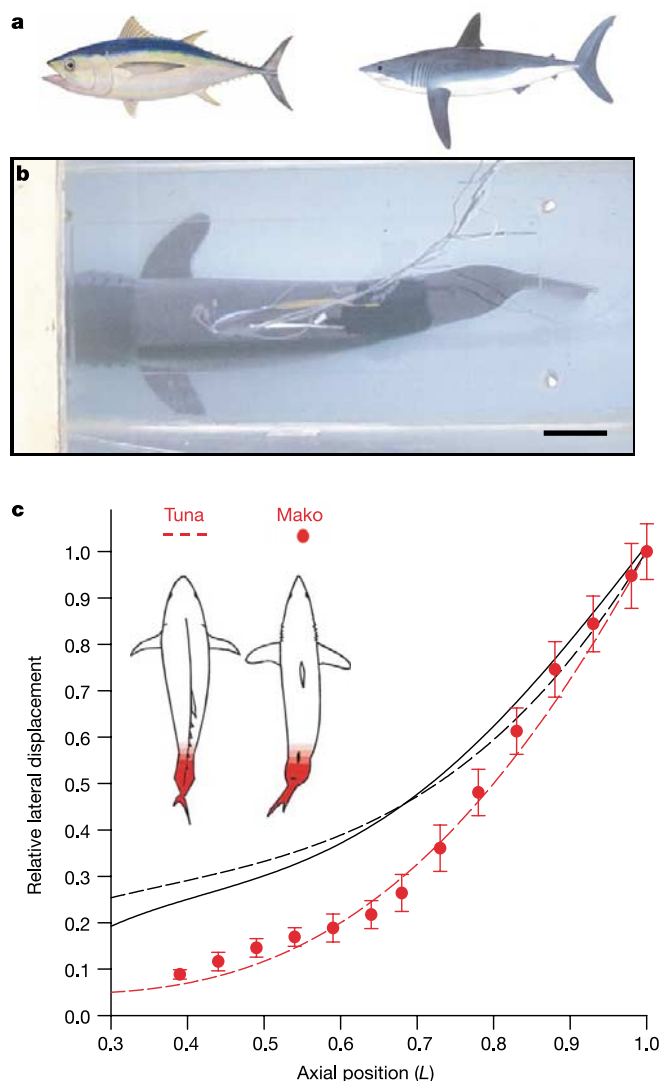


Figure 1 Features of thunniform body shape and patterns of lateral undulation during steady swimming. **a**, Thunniform body shape in lamnid sharks²⁵ (right) and tunas²⁶ (left). Note the highly streamlined fusiform body shape that minimizes pressure drag, stiff, high-aspect-ratio hydrofoil caudal fin that produces thrust by a hydrodynamic-lift mechanism, and dorsoventrally flattened and enlarged caudal keel that decreases drag produced by lateral movement of the peduncle^{6–8}. **b**, Dorsal image of steady swimming *I. oxyrinchus*. Scale bar: 10 cm. **c**, Lateral displacement, relative to the amplitude of lateral motion at the tail tip, versus axial position in the mako shark ($n = 5$; red symbols), tuna (red dashed line²⁷), leopard shark (black solid line¹⁵) and subcarangiform teleosts (black dashed line; modified from ref. 13). This graph emphasizes differences in displacement in the mid-body where most of the variation among swimming modes occurs. Comparison of the mako shark and tuna illustrates the similarity in their swimming mode, where lateral undulations are largely confined to the caudal region, indicated by shading in dorsal outlines. The reduction in lateral motion in the mid-body region afforded by their internal red muscle and modified force-transmission system is evident when comparing the mako shark and tuna to sharks and bony fishes with less-specialized swimming modes such as the subcarangiforms shown here.

during short portions of each contraction cycle the red muscle will be lengthening while the adjacent white fibres are shortening, and vice versa. The loose sheath of connective tissue that surrounds the red muscle mass (see Fig. 3d) may facilitate this shearing. These results are remarkably similar to those described in tunas, where the

deep red muscle shortens and transmits force and displacement to more posterior regions of the body rather than effecting local bending^{4,19}. Thus, the same mechanism enables tunas and lamnid sharks to swim with the relatively stiff-bodied thunniform mode of propulsion; even though the bulk of aerobic muscle is located more

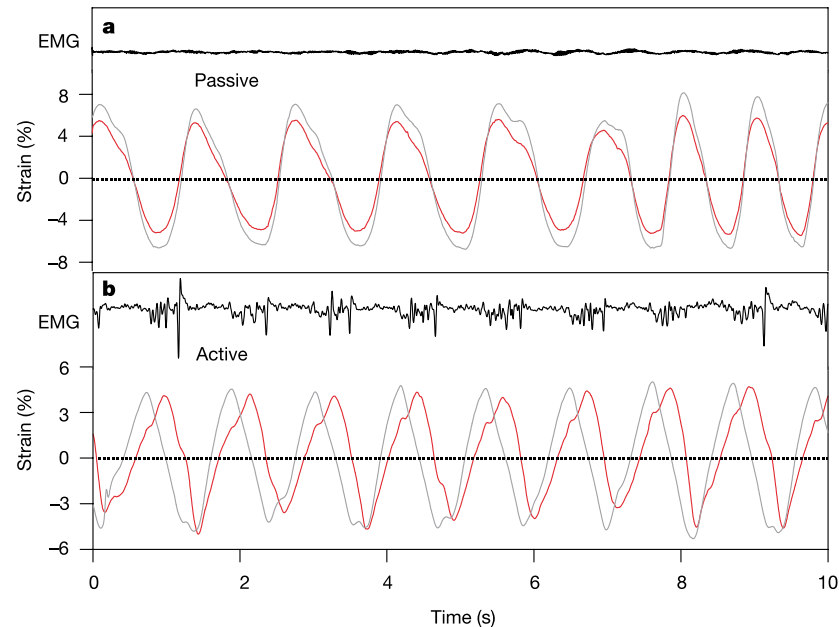


Figure 2 Simultaneous recordings of muscle strain (segment length change/mean length) of red (red trace) and adjacent white (grey trace) muscle during passive simulated swimming movements (**a**) and active steady swimming (0.5 l s^{-1}) (**b**) in the mako shark.

During active swimming, as verified by red muscle activity (EMG trace), shortening in the red muscle is delayed relative to the white muscle and is therefore in phase with lateral motion in more posterior positions.

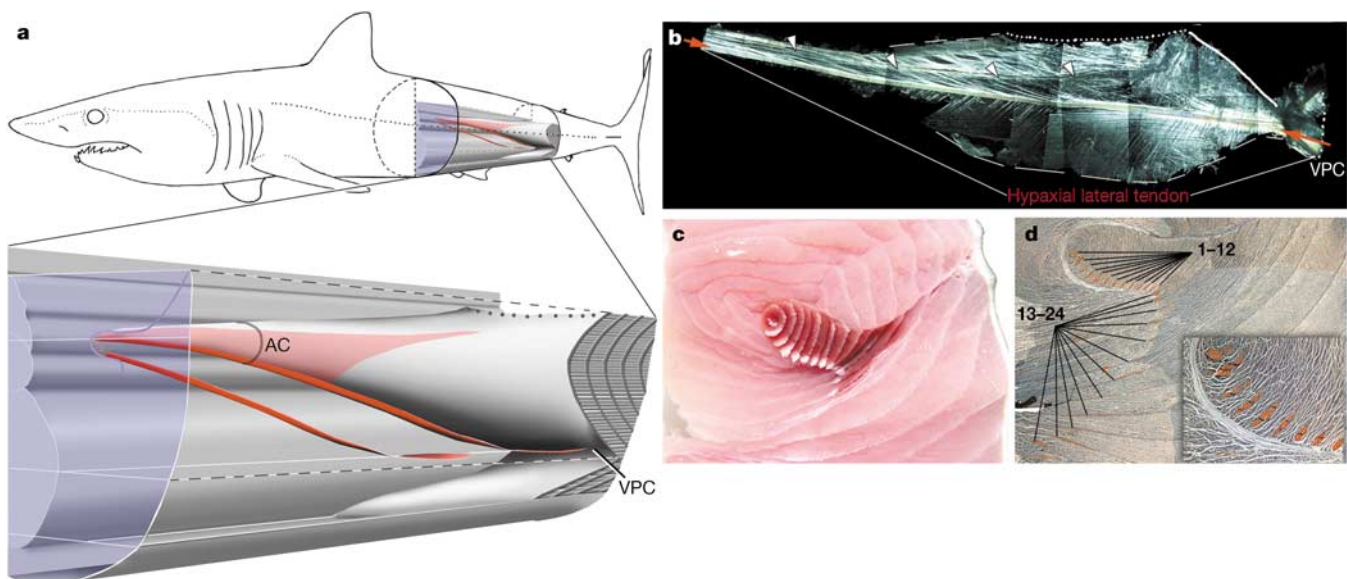


Figure 3 Collagenous architecture of myosepta of *I. oxyrinchus*. **a**, Oblique view focusing on the hypaxial part between 0.54L and 0.74L (coloured inset). The elongated anterior pointing cone (AC) of one myoseptum is shown. It intersects with the red musculature (pale red area) and contains the hypaxial lateral tendon (dark red). The hypaxial lateral tendon extends between the tip of the anterior pointing cone and the ventral posterior cone (VPC). The hypaxial lateral tendon of a more anterior myoseptum is also shown without its myoseptal sheet. The anterior part of this tendon is cut at its intersection with the transverse plane (blue). **b**, Excised area of myosepta between the anterior pointing cone and ventral posterior cone flattened out under polarized light. Pathways of collagenous structures are shown in white. The hypaxial lateral tendon extends between the red arrows. Dashed line, excision line from vertebral axis; thick white line, excision line

from skin; dotted line, excision line from remaining dorsal part of the myosepta, equivalent to the dotted line in **a**; white arrowheads, intersection line of myosepta and loose connective tissue surrounding red muscle. **c**, **d**, Transverse sections of left side. Concentric rings of myosepta indicate nesting anterior pointing cones. **c**, Fresh section, 0.6L, showing red muscle with sections of hypaxial lateral tendons (white) within the red muscle. **d**, Histological section at 0.54L. Red muscle is separated from surrounding white muscle by a sheath of loose connective tissue. Numbers 1–12 indicate anterior portions of 12 hypaxial lateral tendons present in red muscle, whereas numbers 13–24 indicate posterior portions of 12 additional hypaxial lateral tendons present in white muscle, meaning that a single tendon covers 24 segments. The inset shows a detailed view of red muscle and hypaxial lateral tendons (stained orange).

anteriorly than in less derived species, the lateral motion it produces is primarily focused to the caudal region.

Our morphological investigations demonstrate that the anatomical specializations associated with the force-transmission system are also convergent. We used a new combination of techniques to explore the three-dimensional morphology of the tendinous connective tissue linkages (myosepta) that transmit muscular forces to the skin and backbone, and their relationship to the internal red muscle in the mako shark.

In principle, the three-dimensional shape of myotomes and their associated myosepta in mako sharks resembles the regular pattern in gnathostome fishes²⁰, which includes a main anterior cone and a dorsal and ventral posterior cone (see Supplementary Fig. 1 for myoseptal parts of gnathostomes and mako sharks). Additionally, in mako sharks two secondary anterior cones are present at the dorsal and ventral-most part of the myoseptum. The red muscle is situated in the lower part of the main anterior cone (Fig. 3a; see also Supplementary Fig. 1d, e) where its fibres insert into the collagenous myoseptum. In particular, red muscle fibres insert into the anterior half of a myoseptal tendon (Fig. 3a). This tendon runs from the tip of the main anterior cone through the red muscle towards its end within the white muscle at the ventral posterior cone (Fig. 3a, b). It clearly represents the homologue of the hypaxial lateral tendon in gnathostome fishes (Supplementary Fig. 1). In mako sharks, this tendon is extremely prominent and elongated when compared with other fishes. In fact, myoseptal tendons as long and as distinct as those associated with the red muscle in the mako shark have never been reported in any shark species. We measured tendon lengths as long as 0.19L in the posterior region of the body (Supplementary Fig. 1e). The sonomicrometric results suggest that the action of the red muscle is directed posteriorly along the body by 10–17%. The measured tendon lengths accord well with the values predicted from sonomicrometry, suggesting that the hypaxial lateral tendon is responsible for transmitting red muscle forces posteriorly. The prominent tendons of the posterior body are gradually developed along a rostrocaudal gradient from shorter (~0.06L; Supplementary Fig. 1d) and less distinct hypaxial lateral tendons of anterior myosepta to longer tendons in the posterior.

In tunas, distinct and elongated tendons have also been discovered and have been hypothesized to transfer forces from the red muscle to the axial skeleton, and thus provide the anatomical basis for force transmission from the anterior to the caudal region²¹. As in mako sharks, the available length measurements and sonomicrometric data are in good accordance (0.18L experimentally and 0.16L morphologically)²². Although in tunas the primary force-transmitting tendons are in the horizontal septum, in the mako shark, as in other sharks²³, we found the horizontal septum to be reduced in the posterior half of the body. Instead, the primary linkage to the tail appears to be the hypaxial lateral tendons. Interestingly, although the posterior oblique tendons in tunas and hypaxial lateral tendons in mako sharks provide the same function, they have different anatomical origins.

Through distinct evolutionary pathways lamnid sharks and tunas have converged on the same mechanical design principle, that of having internalized red muscle associated with a highly derived force-transmission system, two features that form the basis for their thunniform swimming mode. Our study shows that not only have the physical demands of the external environment sculpted the body shapes of large pelagic cruisers, but also the internal physiology and morphology of their complex locomotor systems has been fine-tuned over the course of their evolution. □

Methods

Shortfin mako sharks (*I. oxyrinchus*, family Lamnidae) ranging in size from 80 to 112 cm L (total body length) were collected by hook and line off the coast of Southern California and transported to the laboratory facilities at Scripps Institution of Oceanography (SIO) in a

transport chamber equipped with circulating aerated sea water. Once at SIO, the sharks were placed into a large 3,000-l swim tunnel for an acclimation period of several hours before experimentation. All procedures in capture, maintenance and experimentation followed the guidelines of the University of California, San Diego Institutional Animal Care and Use Committee.

In vivo muscle dynamics

To examine the dynamics of red and white muscle contractions during swimming, we used electromyography (EMG) and sonomicrometry, a technique for measuring distances in which piezoelectric crystals transmit and receive ultrasonic pulses. Pairs of sonomicrometric crystals were implanted into the deep red and adjacent white muscle to record instantaneous changes in muscle segment length (strain) during active periods of steady swimming as well as during passive, simulated swimming movements induced under anaesthesia. Surgery was performed on anaesthetized individuals partially submerged in a seawater bath according to procedures described previously¹⁵. Crystal pairs were implanted approximately 15 mm apart along the longitudinal axis of the body and the leads were loosely anchored to the skin with surgical sutures. To verify the passive and active states of the red muscle, electrical activity was recorded using pairs of electrodes implanted approximately 2 mm apart directly bisecting the crystal pairs. After surgery, the sharks were placed into the swim tunnel and allowed to recover before data collection. In the recovery period we recorded red and white muscle strain during passive, simulated swimming movements induced by gentle side-to-side motions of the centre of mass that generated body undulation. Additionally, we recorded and analysed 30–50 consecutive tail-beat cycles for each individual while the shark swam steadily at approximately 0.51 s^{-1} . To measure the relative timing of red and white muscle strain (phase shift), a cross-correlation analysis was performed using waveforms containing approximately ten consecutive tail-beat cycles. Mean phase shift presented in the text represents a mean of five individuals. Sonomicrometric and EMG signals were recorded at 500 Hz.

To correlate measurements of local muscle activation and strain with patterns of body bending, five mako sharks were videotaped while swimming against a current of known velocity in the swim tunnel. To synchronize the collection of sonomicrometric, EMG and video recordings, a flashing red diode was recorded in the video sequences and its excitation voltage was recorded with the sonomicrometric and EMG data. Kinematic analysis follows procedures described previously^{15,24}.

Morphology

We used a combination of clearing and staining, microdissection, polarized light microscopy, standard histology and computer-based three-dimensional reconstruction to explore the three-dimensional morphology of the tendinous connective tissue linkages (myosepta). A small body segment (0.54–0.55L) of a formalin-fixed mako specimen (65 cm total length) was prepared for standard histology (paraffin embedding; Azan–Domagk staining, slice thickness of 15 μm). The two remaining parts (0–0.54L and 0.55–1.00L) were carefully skinned, stained for cartilage with Alcian blue 8GX (Merck) and then cleared according to a recently described procedure²⁰. Microdissections on the myoseptal system were carried out using fine microsurgery tools. Myosepta or parts of myosepta from all body regions were removed subsequently. Three-dimensional shape of the myosepta was documented by a camera lucida, and tendon lengths and rostrocaudal extensions of complete myosepta were measured *in situ*. Removed myosepta were photographed under polarized light to visualize the collagen fibre pathways and tendons (Zeiss Stemi 2000C with Fuji X digital camera HC300Z; $1,000 \times 1,450$ pixels). Additionally, the distribution of red muscle and its relation to myoseptal cones along the body was examined. A three-dimensional reconstruction was obtained from histological sections. Major landmarks (vertebrae, neural arches, vertical septum, abdominal cavity, tip of main anterior, secondary anterior and ventral posterior cone, sections of tendons, position of red muscle) were digitized and aligned using SurfDriver 3.5.3. Maxon Cinema 4D (Release 6) was used for choosing an adequate perspective and rendering. The obtained three-dimensional view was edited by Adobe Photoshop (final shading, adding of myoseptal shape).

Received 23 January; accepted 25 February 2004; doi:10.1038/nature02435.

1. Motani, R. Scaling effects in caudal fin propulsion and the speed of ichthyosaurs. *Nature* **415**, 309–312 (2002).
2. Lindsey, C. C. in *Fish Physiology* Vol. VII *Locomotion* (eds Hoar, W. S. & Randall, D. J.) 1–100 (Academic, New York, 1978).
3. Bernal, D., Dickson, K. A., Shadwick, R. E. & Graham, J. B. Analysis of the evolutionary convergence for high performance swimming in lamnid sharks and tunas. *Comp. Biochem. Physiol.* **129**, 695–726 (2001).
4. Katz, S. L., Syme, D. A. & Shadwick, R. E. High speed swimming: enhanced power in yellowfin tuna. *Nature* **410**, 770–771 (2001).
5. Carey, F. G., Teal, J. M., Kanwisher, J. W., Lawson, K. D. & Beckett, J. S. Warm-bodied fish. *Am. Zool.* **11**, 137–143 (1971).
6. Carey, F. G., Casey, J. G., Pratt, H. L., Urquhart, D. & McCosker, J. E. Temperature, heat production and heat exchange in lamnid sharks. *Mem. S. Calif. Acad. Sci.* **9**, 92–108 (1985).
7. Collette, B. B. in *The Physiological Ecology of Tunas* (eds Sharp, G. D. & Dizon, A. E.) 7–40 (Academic, New York, 1978).
8. Graham, J. B. & Dickson, K. A. The evolution of thunniform locomotion and heat conservation in scombrid fishes: new insights based on the morphology of *Allothunnus fallai*. *Zool. J. Linn. Soc.* **129**, 419–466 (2000).
9. Lighthill, M. J. Hydromechanics of aquatic animal propulsion—a survey. *Annu. Rev. Fluid Mech.* **1**, 413–446 (1969).
10. Magnuson, J. J. in *Fish Physiology* Vol. VII *Locomotion* (eds Hoar, W. S. & Randall, D. J.) 239–313 (Academic, New York, 1978).

11. Reif, W. E. & Weishampel, D. B. Anatomy and mechanics of the lunare tail in lamnid sharks. *Zool. Jb. Anat.* **114**, 221–234 (1986).
12. Sfakiotakis, M., Lane, D. M. & Davies, J. B. C. Review of fish swimming modes for aquatic locomotion. *J. Ocean. Eng.* **24**, 237–252 (1999).
13. Altringham, J. D. & Shadwick, R. E. in *Tuna: Physiological Ecology and Evolution* (eds Block, B. A. & Stevens, E. D.) 313–344 (Academic, San Diego, 2001).
14. Webb, P. W. Hydrodynamics and energetics of fish propulsion. *Bull. Fish. Res. Bd Can.* **190**, 1–159 (1975).
15. Donley, J. M. & Shadwick, R. E. Steady swimming muscle dynamics in the leopard shark *Triakis semifasciata*. *J. Exp. Biol.* **206**, 1117–1126 (2003).
16. Coughlin, D. J., Valdes, L. & Rome, L. C. Muscle length changes during swimming in scup: sonomicrometry verifies the anatomical high-speed cine technique. *J. Exp. Biol.* **199**, 459–463 (1996).
17. Katz, S. L., Shadwick, R. E. & Rapoport, H. S. Muscle strain histories in swimming milkfish in steady and sprinting gaits. *J. Exp. Biol.* **202**, 529–541 (1999).
18. Shadwick, R. E., Steffenson, J. F., Katz, S. L. & Knowler, T. Muscle dynamics in fish during steady swimming. *Am. Zool.* **38**, 755–770 (1998).
19. Shadwick, R. E., Katz, S. L., Korsmeyer, K., Knowler, T. & Covell, J. W. Muscle dynamics in skipjack tuna *Katsuwonus pelamis*: timing of red muscle shortening and body curvature during steady swimming. *J. Exp. Biol.* **202**, 2139–2150 (1999).
20. Gemballa, S. et al. Evolutionary transformations of myoseptal tendons in gnathostomes. *Proc. R. Soc. Lond. B* **270**, 1229–1235 (2003).
21. Westneat, M. W., Hoes, W., Pell, C. A. & Wainwright, S. A. The horizontal septum: mechanisms of force transfer in locomotion of scombrid fishes (Scombridae, Perciformes). *J. Morphol.* **217**, 183–204 (1993).
22. Katz, S. L. Design of heterothermic muscle in fish. *J. Exp. Biol.* **205**, 2251–2266 (2002).
23. Gemballa, S., Hagen, K., Röder, K., Rolf, M. & Treiber, K. Structure and evolution of the horizontal septum in vertebrates. *J. Evol. Biol.* **16**, 966–975 (2003).
24. Jayne, B. C. & Lauder, G. V. Speed effects on midline kinematics during steady undulatory swimming in largemouth bass, *Micropterus salmoides*. *J. Exp. Biol.* **198**, 585–602 (1995).
25. Compagno, L. J. V. in *FAO Identification Guide for Fishery Purposes. The Living Marine Resources of the Western Central Pacific* (eds Carpenter, K. E. & Niem, V. H.) 1274–1278 (FAO, Rome, 1998).
26. Joseph, J., Klawe, W. & Murphey, P. *Tuna and Billfish—Fish Without A Country* 1–69 (Inter-American Tropical Tuna Commission, La Jolla, California, 1988).
27. Knowler, T. *Biomechanics of Thunniform Swimming* Thesis, Univ. California, San Diego (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Biewener, J. Gosline, J. B. Graham, S. Vogel and N. Holland for discussion and reviews. Funding was provided by NSF and UC Regents.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.D. (jdonley@ucsd.edu).

Female mating bias results in conflicting sex-specific offspring fitness

Kenneth M. Fedorka* & Timothy A. Mousseau

Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA

* Present address: Department of Genetics, University of Georgia, Athens, Georgia 30602, USA

Indirect-benefit models of sexual selection assert that females gain heritable offspring advantages through a mating bias for males of superior genetic quality. This has generally been tested by associating a simple morphological quality indicator (for example, bird tail length) with offspring viability¹. However, selection acts simultaneously on many characters, limiting the ability to detect significant associations, especially if the simple indicator is weakly correlated to male fitness^{2,3}. Furthermore, recent conceptual developments suggest that the benefits gained from such mating biases may be sex-specific because of sexually antagonistic genes that differentially influence male and female reproductive ability⁴. A more suitable test of the indirect-benefit model would examine associations between an aggregate quality indicator^{1,3} (such as male mating success) and gender-specific

adult fitness components, under the expectation that these components may trade off¹. Here, we show that a father's mating success in the cricket, *Allonemobius socius*, is positively genetically correlated with his son's mating success but negatively with his daughter's reproductive success. This provides empirical evidence that a female mating bias can result in sexually antagonistic offspring fitness.

With the exception of genetic monogamy, the sexes are expected to possess different optimal reproductive phenotypes^{5,6}. Therefore, a conflict may arise because the sexes represent different environments in which homologous fitness-related traits are expressed⁴. Without the added evolution of sex-limited expression (that is, sexual dimorphism), sexual conflict of this nature may constrain the realization of fitness optima⁶. Unfortunately, evidence for this type of conflict is rare. However, recent empirical work in *Drosophila melanogaster* showed that adult relative fitness was negatively correlated between the sexes, even though there was a positive genetic correlation for fitness earlier in development⁴. The authors⁴ further predicted that if these results were applicable outside *Drosophila*, a female mating bias for high-quality mates would produce at best only average-quality daughters.

We addressed this prediction in *A. socius* by examining the genetic correlation between a sire's field-determined mating success and his son's mating success and daughter's reproductive success. We chose these offspring variables on the assumption that they are accurate predictors of next-generation fitness. Previous work suggests that *A. socius* females mate with larger males to gain a larger haemolymph-based nuptial gift. This gift is obtained by chewing a specialized spur on the male hind tibia during copulation^{7,8}. Therefore, the condition of the spur (chewed or unchewed) indicates a male's mating history (successful or unsuccessful): chewing behaviour is rarely uncoupled from sperm transfer. Spur condition also seems to be age-independent because spur categories persist throughout a breeding season that is characterized by little new male recruitment once a cohort emerges⁸. Use of an 'aggregate' indicator, such as mating success, is important because it represents the sum of interacting male traits⁹, overcoming the implicit weakness of the traditional 'simple' indicator approach^{1–3}.

To estimate the sire–offspring correlations, we mated 47 wild-caught sires (25 successful and 22 unsuccessful) to 98 wild-caught females and reared their offspring. Sires were caught mid-breeding season and females were collected as late instar nymphs earlier in the season to ensure their virginity; on average there were two females per sire. Males that had mated in the field were larger than unmated males ($F_{1,45} = 7.25$, $P = 0.0099$) and there was no female size difference between sire groups ($F_{1,96} = 0.01$, $P = 0.9294$). From each female family, two or three sons ($n = 220$ total males) were separately placed into a mating arena with a randomly chosen, unrelated male and female taken from other sire families (see

Table 1 Genetic correlation between a sire's phenotype and offspring variables

Trait	<i>n</i>	Sire mating success		Sire size	
		<i>r_G</i>	<i>P</i>	<i>r_G</i>	<i>P</i>
Son mating success	4.7	0.38	0.0076	0.06	0.7679
Daughter reproductive	6.9	−0.36	0.0164	−0.02	0.9280
Hatching*	545.9	0.01	0.9813	0.07	0.7549
Survivorship*	133.8	0.08	0.5708	0.03	0.8971
Son development*	21.2	−0.02	0.9134	0.01	0.9785
Daughter development*	23.8	0.03	0.8218	0.01	0.5381
Son size*	13.8	0.08	0.5924	0.17	0.9670
Daughter size*	15.0	−0.16	0.2737	0.00	0.9990

* Juvenile viability variable.

Two sire phenotypes were examined, including sire mating success and sire size. *n* represents the average number of individuals per sire family. Bold values remain significant after a sequential Bonferroni. *P*-values corrected within each sire phenotype (mating success and size) and within each fitness stage (that is, adult fitness and juvenile viability).

Methods). The first male to pass a spermatophore was termed 'successful', while the unmated male was termed 'unsuccessful'. To obtain an estimate of male fitness variance, each male was tested on average five times against a different competitor, using a new virgin female for each trial. Thus, a son's mating success was estimated from the proportion of successful mating attempts and a daughter's reproductive success was estimated from her fecundity. In total, 1,067 mating trials were conducted, resulting in 327 inseminated females.

We conducted three additional tests to explore the data further. First, we estimated the genetic correlation between sire mating success and (1) offspring hatching success (ratio of hatchlings to eggs laid); (2) offspring survivorship (ratio of adults to hatchlings); (3) sex-specific offspring development time (days from hatching to adulthood); and (4) sex-specific offspring body size (femur length). This addressed the common indirect-benefit prediction that high-quality sires produce offspring of increased viability. Second, we estimated the genetic correlation between sire size (femur length) and offspring fitness as a contrast to our aggregate-indicator approach. Sire size represents a simple morphological quality indicator: previous work in *A. socius* suggests that females prefer larger males⁸. Third, we estimated the phenotypic correlation matrix between several male reproductive behaviours from the videotaped trials to examine potential mechanisms underlying male mating success. These included courtship duration (time elapsed from courtship onset to copulation); spur chewing duration (the time the female spent chewing the spur—longer chewing results in a larger gift⁷); and spermatophore attachment duration (the time the female kept the spermatophore attached, an estimate of sperm transfer) for each trial.

We found that a sire's field-determined mating success was positively genetically correlated to his son's mating success ($r_G = 0.38$, $P = 0.0076$; Table 1), and negatively genetically correlated to his daughter's reproductive success ($r_G = -0.36$, $P = 0.0164$; Table 1). Thus, successful sires produce successful sons, but unsuccessful daughters. When we calculated average relative offspring fitness, $\bar{w}$ (see statistical analysis) for each sire group, we found that the sons of successful sires were on average 1.7 times fitter than the sons of unsuccessful sires (Fig. 1). In contrast, successful sires produced daughters that were only 0.63 times as fit

as their unsuccessful counterparts. This leads to a 17% greater relative offspring fitness for females mated to successful sires ($(1.7 + 0.63)/2 = 1.17$; after ref. 1). Our data also indicated that offspring sex ratio may vary among sire mating groups (two-tailed t -test: $t_{45} = 1.9$, $P = 0.064$). Accordingly, we recalculated $\bar{w}$ after correcting for sex ratio and found that offspring of successful sires were 24% more fit than offspring of unsuccessful sires. Assuming that these fitness estimates are accurate approximations of lifetime success, then female preference for high-quality males can still be maintained in the face of sexually antagonistic offspring fitness.

To determine whether the bias in daughter fecundity was due to a bias in mate quality, we examined the relative proportion of 'successful sons' (that is, sons of successful sires) mated to successful and unsuccessful daughters. We found that 60.6% (94/155) of the 'unsuccessful daughters' and 66.3% (114/172) of the 'successful daughters' were mated to successful sons. These groups were not significantly different ($\chi^2 = 1.12$, degrees of freedom, d.f. = 1, $n = 327$, $P = 0.2904$). This suggested that the difference in daughter fecundity between sire groups was not due to a bias in the quality of their mates.

Analysis of the mating trials indicated that the sons' reproductive phenotypes were significantly repeatable (Table 2; see Methods). Because these estimates are based on the intra-class correlation, they provide a far-upper-bound estimate of the underlying additive genetic variance. A son's femur length was positively correlated to his mating success ($r_P = 0.24$, $n = 220$, $P = 0.0003$) and gift size ($r_P = 0.33$, $n = 220$, $P < 0.0001$). Therefore, we examined the phenotypic correlation matrix between reproductive phenotypes using a partial correlation to adjust for the influence of body size. These results indicated that the more successful sons acquired mates more quickly (mating success and courtship duration correlation: $r_P = -0.27$, $n = 220$, $P = 0.0015$).

We found no significant genetic correlation between sire success and offspring hatching success, survivorship, son/daughter development time or body size (Table 1, all $P > 0.27$). This implies that the female mating bias for successful sires does not increase her offspring's viability. Furthermore, these data suggest that confounding non-genetic parental effects are minimal or non-existent, given that sire 'quality' had no effect on pre-adult components of fitness, which is where we would expect these effects to occur^{10–13}. In addition, we found no significant genetic correlation between sire size (simple quality indicator) and any offspring fitness variable (Table 1), suggesting that an aggregate-indicator of male quality may be more revealing.

Our data support the hypothesis that mating biases for high-quality males can result in conflicting sex-specific offspring fitness returns. These data provide general support for the intersexual genetic conflict model of evolution^{4,6}, suggest an alternative approach in testing indirect-benefit models (for example, aggregate quality indicators coupled with sex-specific adult fitness estimates), and represent one of a few studies addressing the genetic relationship between father and son mating success.

The inequality of sex-specific fitness returns as presented here has several important consequences. First, owing to the transient nature of the sexual environment in which the underlying fitness loci are expressed, ample antagonistic fitness variation can be maintained through opposing gender-specific selection^{4,6}. Second, opposing selection will cause the realization of gender-specific fitness optima

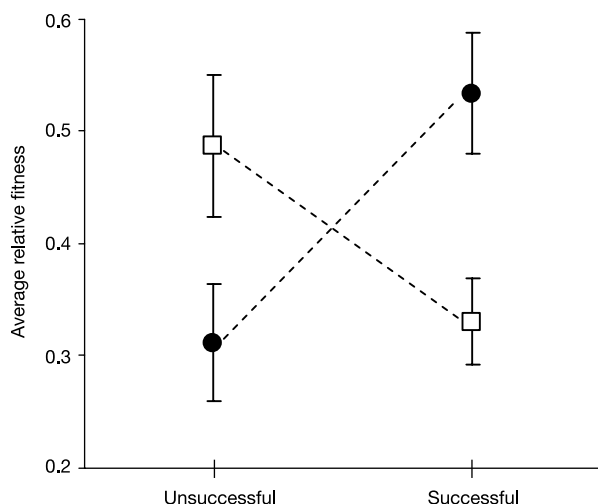


Figure 1 Female mating bias resulted in conflicting sex-specific offspring fitness. Successful sires produced sons with a higher mating success (circles), but daughters with a lower reproductive success (squares) when compared to unsuccessful sires (means $\pm$ s.e. are based on sire family means). The dashed line connects the same sex.

Table 2 Trait repeatability

Trait	d.f.	F	R	P <
Mating success	219, 1065	3.13	0.47	0.0001
Courtship	194, 568	1.51	0.15	0.0004
Gift size	144, 343	1.81	0.26	0.0001
Spermatophore	142, 326	1.28	0.11	0.0580

to be constrained⁴. Third, future sexual selection models should acknowledge that a mating bias for indirect-benefits or 'good genes' can result in conflicting sex-specific offspring fitness⁴. □

Methods

Organismal maintenance

Crickets were maintained in plastic cages (10 × 10 × 8 cm) containing ground cat food and a carrot slice (provided ad libitum), dampened cheesecloth (water source and oviposition material) and strips of brown paper towel for cover. The food, carrot and paper towel were replaced every two days. Cages were kept in a constant environment at 28 °C and a 12:12 light–dark photoperiod provided by a Percival incubator. The age of laboratory-reared experimental crickets was 10 ± 1 days post-eclosion (final adult moult).

Mating trials

Mating arenas were made from 100-mm Petri dishes lined with filter paper. Males from both successful and unsuccessful sires were randomly chosen for each mating trial to provide an equal sire-group representation. Females were also randomly chosen from either sire group. To help control for the potential influence of mating experience, males were tested against other males who had shared an equal number of previous trials (that is, 0, 1, 2, and so on). Each male was mated once per day. After mating, females were isolated to oviposit until their death. All trials were videotaped.

Statistical analysis

Repeatabilities (R) were estimated by calculating the intra-class correlation obtained from a one-way ANOVA (that is, $R = (MS_{\text{among}} - MS_{\text{error}}) / [MS_{\text{among}} + (N - 1)MS_{\text{error}}]$, where MS is the mean square estimate and N is the number of repeated measures per individual)¹⁴. Genetic correlation coefficients were based on sire family means ($n = 47$). Because correlation estimates were cross-generational (that is, sire–offspring), estimates of trait (co)variance were based on different individuals within each sire family (as opposed to the same individuals), thereby providing a relatively unbiased estimator of the additive genetic correlation. Tests of significance were based on correlation-coefficient critical values derived from standard statistical tables¹⁵.

For each sex, $\bar{w}$ among sire groups was calculated as:

$$\bar{w}_{jg} = \left(\sum_i \frac{w_{ji}}{w_{j\max}} \right) / n_{jg} \quad (1)$$

where w is the average offspring fitness of sex j for the i th sire; $w_{\max}$ represents the maximum observed fitness value; g represents the sire group (successful or unsuccessful); and n is the number of sires. Thus, four estimates of average relative fitness were calculated including the sons and daughters of successful sires, as well as the sons and daughters of unsuccessful sires. To correct for potential differences in sex ratio between sire groups, we recalculated $\bar{w}$ by first multiplying w_{ji} by a_{ji} , where a is the offspring abundance of sex j for the i th sire. Each new male and female estimate of $\bar{w}$ was then divided by their respective population sex ratios (male:female is 0.89 and female:male is 1.12), to adjust for the difference in the fisherian value between the sexes; that is, because fewer males existed, they possessed a higher breeding value. All analyses were performed using SAS v8.

Received 4 February; accepted 9 March 2004; doi:10.1038/nature02492.

- Kokko, H. Fisherian and 'good genes' benefits of mate choice: how (not) to distinguish between them. *Ecol. Lett.* **4**, 322–326 (2001).
- Boake, C. R. B. Genetic consequences of mate choice: a quantitative genetic method for testing sexual selection theory. *Science* **227**, 1061–1063 (1985).
- Heisler, I. L. Quantitative genetic models of female choice based on "arbitrary" male characters. *Heredity* **55**, 187–198 (1985).
- Chippindale, A. K., Gibson, J. R. & Rice, W. R. Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 1671–1675 (2001).
- Rice, W. R. Dangerous liaisons. *Proc. Natl Acad. Sci. USA* **97**, 12953–12955 (2000).
- Chapman, T., Arnqvist, G., Bangham, J. & Rowe, L. Sexual conflict. *Trends Ecol. Evol.* **18**, 41–47 (2003).
- Fedorka, K. M. & Mousseau, T. A. Tibial spur feeding in ground crickets: larger males contribute larger gifts (Orthoptera: Gryllidae). *Florida Entomol.* **85**, 317–323 (2002).
- Fedorka, K. M. & Mousseau, T. A. Nuptial gifts and the evolution of male body size. *Evolution* **56**, 590–596 (2002).
- Wedell, N. & Tregenza, T. Successful fathers sire successful sons. *Evolution* **53**, 620–625 (1999).
- Mousseau, T. A. & Dingle, H. Maternal effects in insect life histories. *Annu. Rev. Entomol.* **36**, 511–534 (1991).
- Mousseau, T. A. & Dingle, H. In *The Unity of Evolutionary Biology* (ed. Dudley, E. C.) 745–761 (Dioscorides, Portland, Oregon, 1991).
- Roff, D. A. *Evolutionary Quantitative Genetics* (Chapman & Hall, New York, 1997).
- Montalvo, A. M. & Shaw, R. G. Quantitative genetics of sequential life history and juvenile traits in the partially selfing perennial, *Aquilegia caerulea*. *Evolution* **48**, 828–841 (1994).
- Zar, J. H. *Biostatistical Analysis* 2nd edn, 324 (Prentice Hall, Englewood Cliffs, New Jersey, 1984).
- Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* 632–637 (Sinauer Associates, Sunderland, 1998).

Acknowledgements We thank D. Promislow, P. Mack and J. Burger for comments on previous manuscript versions. We also thank N. Leung and A. Penn for assisting with cricket maintenance and data collection. This work was supported by a NSF Doctoral Dissertation Improvement Grant to K.M.F. and a NSF grant to T.A.M.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.M.F. (fedorka@uga.edu).

Naturalistic experience transforms sensory maps in the adult cortex of caged animals

Daniel B. Polley^{1*}, Eugen Kvašňák¹ & Ron D. Frostig^{1,2,3}

¹Department of Neurobiology and Behavior, ²Department of Biomedical Engineering and ³The Center for the Neurobiology of Learning and Memory, University of California, Irvine, California 92697-4550, USA

* Present address: Keck Center for Integrative Neuroscience, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, California 94143-0732, USA

Much of what is known about the functional organization and plasticity of adult sensory cortex is derived from animals housed in standard laboratory cages^{1,2}. Here we report that the transfer of adult rats reared in standard laboratory cages to a naturalistic habitat modifies the functional and morphological organization of the facial whisker representation in the somatosensory 'barrel' cortex. Cortical whisker representations, visualized with repeated intrinsic signal optical imaging in the same animals, contracted by 46% after four to six weeks of exposure to the naturalistic habitat. Acute, multi-site extracellular recordings demonstrated suppressed evoked neuronal responses and smaller, sharper constituent receptive fields in the upper cortical layers (II/III), but not in the thalamic recipient layer (IV), of rats with naturalistic experience. Morphological plasticity of the layer IV barrel field was observed, but on a substantially smaller scale than the functional plasticity. Thus, transferring animals to an environment that promotes the expression of natural, innate behaviours induces a large-scale functional refinement of cortical sensory maps.

In a previous study, we demonstrated that natural behaviours could have a powerful impact on the expression of plasticity in whisker-deprived adult rats, given a brief opportunity for natural whisker use outside the standard home cage³. We found that a few minutes of natural whisker use per week induces a profound contraction of the remaining (spared) whisker's cortical representation, and a decrease of its peak amplitude in contrast to the expansion and increase in peak amplitude of the remaining whisker representation that is commonly observed in rats that remain exclusively in their home cage^{3–8}. In this study, we examined whether the effects of natural whisker use are limited to the cortex of whisker-deprived rats, or if similar plasticity can be observed in the barrel cortex of non-deprived adult rats, given greater opportunity for natural whisker use outside the home cage. Rather than simply elevate sensory activity through a traditional 'enriched environment' (typically, a larger cage filled with various toys, a running wheel and conspecifics), we used a new type of environment, the naturalistic habitat (NH), which promotes innate sensorimotor behaviours such as subterranean tunnelling, foraging and three-dimensional navigation, in addition to interactions with conspecifics (Supplementary Fig. S1).

We also compared functional plasticity induced through naturalistic experience to potential changes in the anatomical architecture of the cortex. The barrel cortex is uniquely suited to address this question, given the isomorphic representation of each whisker on the snout with an isolated cluster of granular cells in layer IV of the contralateral posteromedial barrel subfield (PMBSF), called a barrel⁹. Although the gross morphology of the layer IV barrel map can be markedly altered, this effect is strictly limited to the first week of postnatal life^{10,11}. The absence of anatomical reorganization in the adult PMBSF seems to be at odds with reports that experience in traditional enriched environments can produce a host of anatomical changes in the adult neocortex^{12–14}. This discrepancy may be caused by a difference in methodology; the critical period for barrel field plasticity has been defined from methods that create imbalances in the level of sensory activity through denervation or deafferentation of the whisker pad, whereas anatomical plasticity in enriched

environments is thought to result from enhancing the amount and complexity of sensory activity.

The area and evoked peak amplitude of individual functional whisker representations were measured with chronic intrinsic signal optical imaging (ISI), before and after rats were exposed to the standard cage (SC) or NH. The area and overlap of individual functional whisker representations did not change systematically in the four to six week interval between imaging sessions in SC rats (Fig. 1a–d). In contrast, after four to six weeks of exposure to the NH, functional whisker representations were extensively contracted and less overlapping in NH rats (Fig. 1e–h). A quantitative analysis of functional whisker representations assessed at multiple activity thresholds demonstrated that the average area of a single whisker's functional representation did not differ before assignment to the NH or SC ($F_{1,19} = 0.07$; $P > 0.05$; Fig. 2a). In the second imaging session the area of a single whisker's representation contracted by an average of 46% in NH rats ($F_{1,12} = 14.58$; $P < 0.005$), but did not change significantly for SC rats ($F_{1,7} = 0.12$; $P > 0.05$), yielding a significant interaction between treatment groups and imaging session ($F_{1,19} = 7.22$; $P < 0.05$; Fig. 2b). In addition to the areal contraction, we observed a significant reduction in the whisker-evoked peak amplitude in NH rats ($-19.77 \pm 6.0\%$, $t = 3.01$; $P < 0.05$), but not in SC rats ($-2.07 \pm 4.26\%$, $t = 0.7$; $P > 0.05$; Fig. 2c), between the first and second imaging session. To determine whether the decrease in the area of a single whisker's representation (measured at absolute optical response thresholds) was attributable to a decrease in the overall amplitude of the whisker-evoked response, areal extent measurements were also made after activity levels were normalized to the peak response. The area at half height measured with this approach did not change between the first and second imaging session in SC rats (3.24 ± 0.67 versus 3.17 ± 0.42 mm²), but contracted by 46.3% between the first and second imaging sessions in NH rats (3.2 ± 0.42 versus 1.72 ± 0.31 mm²), demonstrating that a single whisker's representation contracted within the functional whisker map (Fig. 2d). Additional analyses of the ISI data demonstrated that the contraction could not be accounted for by a change in baseline signal

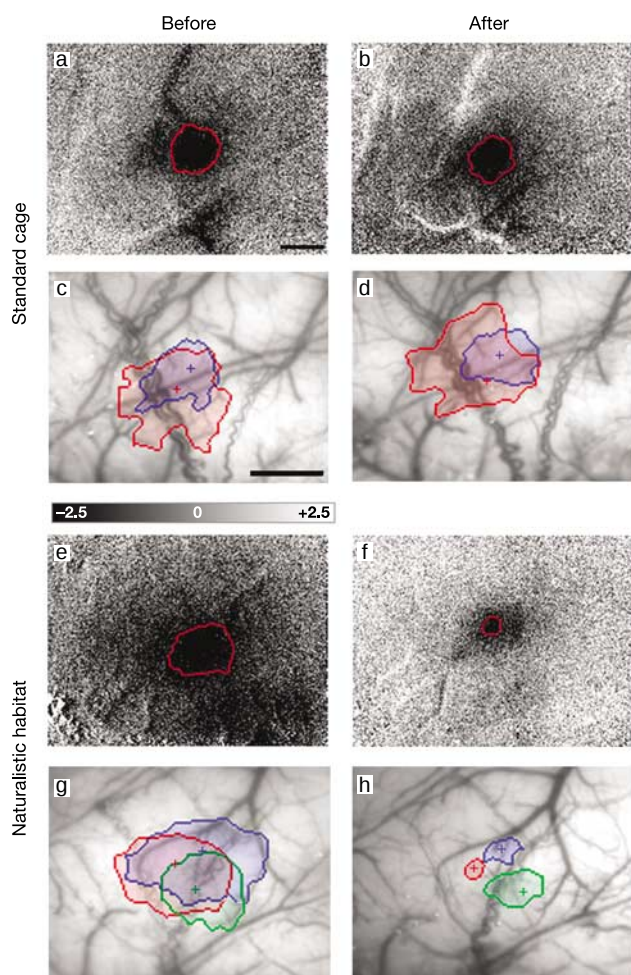


Figure 1 Chronic optical imaging of functional whisker representations through the intact skull. For all panels, Colour boundaries demarcate cortical areas with intrinsic signal activity $\geq 2.5 \times 10^{-4}$. **a, b, e, f**, Ratio images of the C2 whisker representation collected before (**a** and **e**) and after (**b** and **f**) exposure to SC (**a** and **b**) or NH (**e** and **f**). Greyscale bar indicates intrinsic signal strength $\times 10^{-4}$. Black and white streaks correspond to large surface blood vessels. Scale bar in **a**, which also applies to **b, e** and **f**, = 1 mm. **c, d, g, h**, Functional representations for whisker C1 (blue), C2 (red) and B2 (green; in **g** and **h**), obtained from two different SC (**c** and **d**) or NH (**g** and **h**) rats, superimposed onto images of the cortical surface as seen through the thinned skull before (**c** and **g**) and after (**d** and **h**) exposure. Functional peaks for each whisker are indicated by a cross. Scale bar in **c**, which also applies to **d, g** and **h**, = 1 mm.

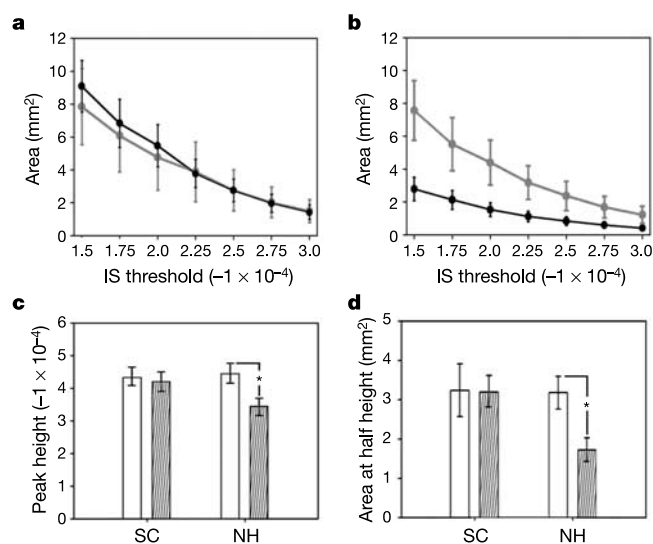


Figure 2 Plasticity of functional whisker representations. **a, b**, Areal extent measured across seven activity thresholds of a single whisker's functional representation, imaged before (**a**) and after (**b**) exposure to NH (black line) and SC (grey line) conditions. IS, intrinsic signal. **c**, Peak height and **d**, area at half height are shown before (open bars) and after (patterned bars) exposure. Asterisks indicate a statistically significant difference ($P < 0.05$) between groups using a paired *t*-test.

strength, or a shift in the temporal profile of the intrinsic signal response (see Supplementary Information). Finally, the contraction of the whisker representation was confirmed by post-imaging single unit recordings (Supplementary Fig. S2).

Next, using simultaneous recordings from an eight-electrode array, we measured receptive field size and whisker-evoked response strength in 349 single and multiunit clusters—located in layer II/III (NH, $n = 90$; SC, $n = 81$) and layer IV (NH, $n = 112$; SC, $n = 66$) from five SC and five NH rats—to determine how the plasticity of cortical population responses observed with ISI corresponded to plasticity in neuronal response properties (Supplementary Fig. S3). Receptive field size measured in layer IV was highly similar in SC and NH rats (median receptive field size is four whiskers for both NH and SC neurons; two-sample Kolmogorov–Smirnov test, $P > 0.05$; Fig. 3b). However, receptive field size in layer II/III neurons was significantly smaller in NH rats compared with SC rats (median receptive field size is two whiskers in NH rats and six whiskers in SC rats; two-sample Kolmogorov–Smirnov test, $P < 0.005$; Fig. 3a). This effect provides a plausible neurophysiological substrate for the contraction of a single whisker's functional representation reported with optical imaging, but it does not indicate whether the reduction in a whisker representation's peak amplitude corresponded to a reduction in neuronal firing rate at the centre of the receptive field and/or the surround. The response strength of layer II/III neurons showed a significant suppression to stimulation of the principal whisker (16.13 ± 1.72 versus 22.87 ± 2.3 spikes per second) and adjacent whisker (5.3 ± 0.77 versus 8.4 ± 1.23 spikes per second) in NH compared with SC rats, respectively ($P < 0.05$; Mann–Whitney U -test). In contrast, there was no difference in the principal whisker or adjacent whisker response strength in layer IV between NH and SC rats (see univariate response distributions in Fig. 3c, d).

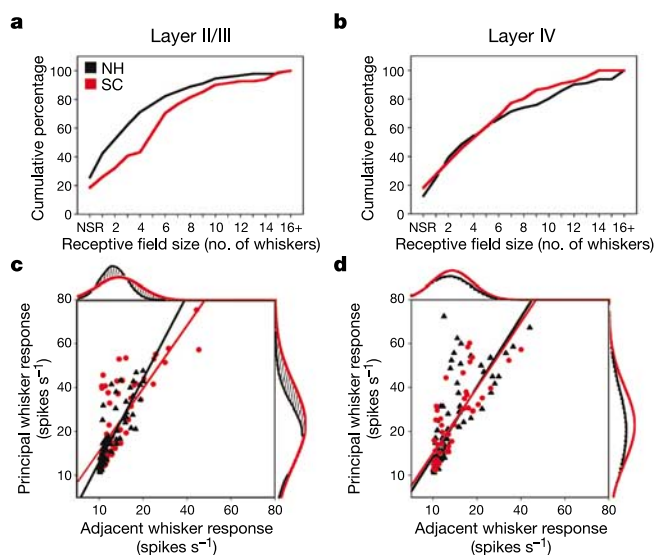


Figure 3 Receptive field plasticity in layer II/III but not IV. Red and black lines represent SC and NH values, respectively. **a**, **b**, Cumulative percentage graphs represent the distribution of receptive field sizes for neurons recorded in layer II/III (**a**) and IV (**b**). Neurons not significantly responsive to stimulation of any whiskers were identified as NSR. **c**, **d**, Scatter plots of unit response magnitude to principal and adjacent whisker stimulation for neurons in layer II/III (**c**) and IV (**d**). The principal whisker (right) and adjacent whisker (top) univariate response distributions are shown. Stippled regions indicate significant ($P < 0.05$) differences in the principal-whisker- and adjacent-whisker-evoked response magnitude.

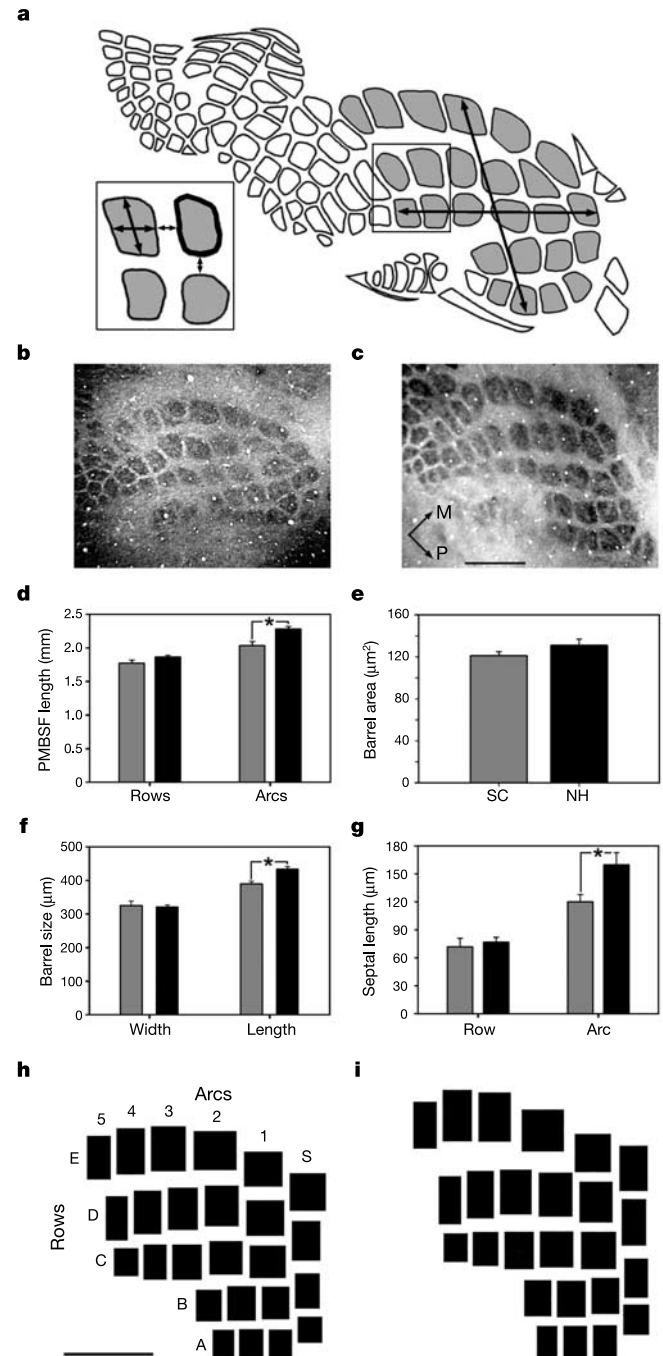


Figure 4 Characterizing morphological plasticity of the layer IV barrel field. **a**, Illustration of the PMBSF (grey) seen in a tangential section through layer IV. Arrows indicate the axis for measuring the length of the PMBSF along arcs (vertical arrow) and rows (horizontal arrow). Inset: axis of measurement for barrel length (vertical arrow), width (horizontal arrow), area (thick line circumscribing the barrel) and the length of septa separating barrels along a row and arc. **b**, **c**, Cytochrome oxidase reactivity in flattened sections through layer IV of the PMBSF in SC (**b**) and NH (**c**) rats. Arrows indicate medial (M) and posterior (P). Scale bar in **c**, which also applies to **b**, = 1 mm. **d**, PMBSF length across rows and arcs. **e**, Individual barrel area. **f**, Length and width of individual barrels. **g**, Length of septa separating pairs of adjacent barrels along a row or an arc. Asterisks indicate statistically significant differences between NH (black) and SC (grey) groups ($P < 0.05$) using an unpaired t -test. **h**, **i**, Mean length and width of each individual barrel as well as the mean length of the septa separating each adjacent barrel along a row and arc are drawn to scale to reconstruct the average cytochrome oxidase map for SC (**h**) and NH (**i**) rats. Scale bar = 1 mm.

Interestingly, the suppression was not expressed uniformly across the receptive field. The principal whisker response bias (a ratio of principal whisker/adjacent whisker response strength) was significantly greater in layer II/III neurons from NH versus SC rats (3.8 ± 0.86 versus 2.48 ± 1.0 ; $P < 0.05$, Mann–Whitney U test) but was not different between layer IV neurons. Thus, as illustrated by the difference in the linear regression slopes in Fig. 3c, the response to a stimulus placed in the surround of the receptive field was more strongly suppressed than the response to a stimulus placed at the centre of the receptive field, tantamount to a sharpening of the receptive field rather than a general reduction in response strength. Finally, we compared the spontaneous firing rate (measured immediately before stimulus onset) in SC and NH rats for layer II/III (SC versus NH: 2.96 ± 0.27 versus 3.03 ± 0.27 spikes per second) and layer IV (3.46 ± 0.34 versus 2.98 ± 0.24 spikes per second), but did not find any significant differences between treatment group ($F_{1,347} = 0.73$; $P > 0.05$) or layer ($F_{1,347} = 0.56$; $P > 0.05$). Collectively, these results demonstrate that plasticity is differentially expressed across the receptive field and is specific to evoked, not spontaneous, neuronal responses in the upper cortical layers, and therefore, cannot be attributed to an increased sensitivity to anaesthesia, decreased cortical excitability or a nonspecific adaptation to simple whisker stimuli.

To determine whether the plasticity of the functional whisker map was accompanied by changes in the morphological whisker map, we processed cortical sections from NH and SC rats for cytochrome oxidase reactivity (an activity-dependent enzyme with an expression pattern that co-localizes with the centre of the cytoarchitecturally defined barrels within the PMBSF of the rat¹⁵). Examination of cytochrome oxidase maps from a representative SC (Fig. 4b) and NH rat (Fig. 4c) demonstrates that although the general morphology of the PMBSF is quite similar between groups, the length of the PMBSF in NH rats (2.28 ± 0.04 mm) is modestly, but significantly, longer along the arc axis than in SC rats (2.03 ± 0.07 mm; $t = 3.31$; $P < 0.01$), but was not different along the row axis (SC versus NH: 1.77 ± 0.05 versus 1.86 ± 0.03 mm; $t = 1.69$; $P > 0.05$; Fig. 4d). The constituent barrels were not greater in area (SC versus NH: 121.0 ± 6.23 versus 131.11 ± 3.51 μm^2 ; $t = 1.41$; $P > 0.05$; Fig. 4e) or width (SC versus NH: 325.0 ± 13.1 versus 321.11 ± 6.33 μm ; $t = 2.58$; $P > 0.05$; Fig. 4f), but the length was slightly larger in NH rats (434.44 ± 7.29 μm) compared with SC rats (389.0 ± 9.36 μm ; $t = 3.77$; $P < 0.05$; Fig. 4f). Similarly, the average length of the septa separating barrels along an arc was significantly longer in NH rats (158.89 ± 12.85 μm) compared with SC rats (123.0 ± 8.44 μm ; $t = 2.38$; $P < 0.05$; Fig. 4g), but no differences were observed in the length of the septa along rows (SC versus NH: 72.0 ± 2.91 versus 76.67 ± 5.27 μm ; $t = 0.78$; $P > 0.05$; Fig. 4g).

Further analysis showed that the increase in barrel length was small and homogenous across the PMBSF, but the increase in septal length was spatially restricted and comparatively larger (Fig. 4h, i). For each animal, the PMBSF geometry was reconstructed through measurements of the length and width of all individual barrels, as well as the length of the septa separating each pair of adjacent barrels (Supplementary Fig. S4). Comparison of the averaged PMBSF maps for SC and NH groups demonstrated an increase in barrel length of NH rats in 24 of 25 barrels measured, but by only 10.2% on average (Fig. 4h, i). In contrast, the septal expansion was proportionately much larger and expressed almost exclusively in the anteromedial corner of the PMBSF. In NH rats compared with SC rats, respectively, we observed significant increases in the length of the septa separating barrels D3 and E3 (307 ± 0.03 versus 177 ± 0.02 μm ; $t = 3.42$; $P < 0.01$), D4 and E4 (324 ± 0.06 versus 177 ± 0.03 μm ; $t = 2.8$; $P < 0.05$) and D5 and E5 (302 ± 0.04 versus 182 ± 0.01 μm ; $t = 3.09$; $P < 0.05$). Additionally, we found no difference in the thickness of the cortex (measured in coronal sections through the PMBSF) in NH rats (1.59 ± 0.04 mm) compared to SC rats

(1.63 ± 0.05 mm; unpaired t -test, $t = 0.54$; $P > 0.05$). Further research is needed to determine whether the source of the morphological changes observed in layer IV originate from changes in the thalamocortical connectivity patterns or from changes that originate within layer IV itself and are thus independent of the thalamocortical input.

We found that individual whisker representations contract to nearly half of their original area after one month in the NH. The magnitude of this effect is surprisingly large given the age of the animals and the absence of an imbalance of sensory inputs at the periphery (for example, whisker trimming or follicle lesion) which would strongly drive competitive plasticity mechanisms. The functional contraction and response suppression that was reported after prolonged exposure to the NH might have been consolidated from the rapid attenuation of sensory-evoked responses in the barrel cortex, known to occur during bouts of active whisking in awake rats^{16,17}. In support of this finding, it has recently been shown that the peak amplitude and spatial spread of whisker-evoked activity in the barrel cortex are strongly reduced and receptive fields are sharpened during acute stimulation of the reticular formation, a technique that imitates the physiological conditions of natural arousal^{17,18}. Furthermore, increasing sensory activity through early exposure to enriched environments^{19,20} or by continuous whisker stimulation²¹, increases response selectivity in the cortex, perhaps through upregulating inhibitory intracortical synapses²² or increasing levels of neurotrophins such as brain-derived neurotrophic factor²³. The laminar specificity for these changes reinforces the concept that supragranular layers of the barrel cortex remain plastic for a longer period than the granular layer^{5,7,24}, and that these changes are probably mediated through plasticity of intracortical circuits^{25,26}. In our previous study, we reported that a few minutes per week of natural whisker use outside the home cage induced a contraction of whisker representations and a reduction of their peak amplitude in the barrel cortex of adult rats with partially removed whisker arrays, but had no effect on the area of a single whisker's representation in rats with fully intact whisker arrays³. Our data demonstrate that this type of plasticity can also occur in layer II/III of non-deprived adult rats that engage in natural behaviours, suggesting that smaller, more metabolically efficient whisker representations and smaller, sharper receptive fields (an apparent refinement of cortical maps) are a characteristic feature of cortical organization in rats that are allowed sufficient opportunity to express innate behaviours. □

Methods

Housing conditions

Male Sprague–Dawley rats were reared in standard cages until they were at least 90 days old. Rats were then randomly assigned to live in NH or modified SC conditions. The NH was created from a circular steel tank, 2 m in diameter and 1 m deep, filled to approximately 70% capacity with packed, sterile topsoil. An extensive labyrinth of tunnels containing objects with a wide variety of textures and shapes was built on top of the packed soil. Eight to ten rats were simultaneously placed in the NH and remained there continuously for 28 to 42 days. Rats assigned to the SC were housed in standard individual cages ($33 \times 22 \times 15$ cm) for 28 to 39 days. Standard cages were filled with packed soil from the NH and placed side-by-side less than a meter away from the NH to make the level of non-somatosensory stimulus modalities more similar to NH rats.

Surgical procedures

Rats were anaesthetized with sodium pentobarbital (50 mg kg^{-1} followed by supplements as needed, approximately 15 mg kg^{-1}). In optical imaging experiments, the skull overlying the left barrel cortex was thinned and enclosed with a petroleum jelly well, filled with saline to create a transparent, non-invasive cortical window. In electrophysiological recording experiments, a craniotomy was performed over the left barrel cortex and the dura was reflected.

Optical imaging

A detailed description of the chronic ISI procedure has been described previously²⁷. Chronic high-resolution images were obtained from adult male rats (375–600 g) before and after assignment to NH ($n = 13$) or SC ($n = 8$) housing conditions. Images were obtained under stabilized 630 nm light with a slow-scan CCD camera (Quantix, Roper

Scientific), defocused 500 μm beneath the cortical surface. Sixty-four single trials were summed and divided into eleven 500 ms frames for image visualization and analysis. After 1 s of prestimulus data acquisition, an individual whisker was displaced 1.9° rostrocaudally at 5 Hz for 1 s.

Electrophysiology

Extracellular recordings of 449 single and multiple units were obtained from seven NH and seven SC rats. In some cases ($n = 100$ units; two rats from each group), layer II/III recordings were obtained only to verify correspondence with optical imaging data (see Supplementary Fig. S2). In all other cases ($n = 349$), individual whisker representations were visualized immediately before electrophysiological recordings with ISI to guide electrode placement into specific regions of the barrel field (Supplementary Fig. S3). Neural activity was filtered, amplified and sorted simultaneously from an array of eight low impedance (1–2 m Ω) tungsten microelectrodes (Microprobe) using a multi-channel acquisition and online spike sorting system (Alpha Omega). Recordings were obtained from 7–8 locations in layer II/III and layer IV in each rat. Often, more than one single unit waveform was isolated from a single electrode. Electrode depth was matched to cortical laminae by post-mortem histology.

Cytochrome oxidase histochemistry

A separate set of 19 rats (NH, $n = 9$; SC, $n = 10$) were perfused intracardially and the left hemisphere was sectioned tangentially. In some cases (NH, $n = 8$; SC, $n = 7$), the right hemisphere was also removed and sectioned in the coronal plane. All sections were stained for cytochrome oxidase reactivity using a standard rat protocol¹⁵.

Analysis of optical imaging data

A detailed account of areal extent analysis procedures can be found elsewhere²⁸. Briefly, an intratrial division analysis designed to focus on the initial decrease in deoxyhaemoglobin levels was applied to each square 34 μm pixel, within the 191 $\times$ 144 pixel image. Ratio images were created by dividing poststimulus intrinsic signal activity (collected 0.5 s up to 1.5 s poststimulus onset) by prestimulus intrinsic signal activity (collected during the 500 ms immediately preceding stimulus onset). Ratio images were gaussian filtered (half width = 5) and the areal extent of each whisker's functional representation was quantified at seven threshold levels above the median prestimulus activity level (from 1.5×10^{-4} to 3.0×10^{-4} , in 0.25×10^{-4} increments). In all but four rats, it was possible to image the representation of several single whiskers (whiskers C1 and C2 in 14 rats; C1, C2 and B2 in 3 rats). Whiskers were stimulated individually in all cases. Areal extent values obtained for each whisker were averaged so that each rat contributed a single value for all measures. All values are expressed as mean $\pm$ standard error. The peak of a functional whisker representation was defined as the single pixel with the greatest post-/pre-ratio value. Peak amplitude is calculated as the peak intrinsic signal response minus median prestimulus intrinsic signal response. It has recently been demonstrated that the peak of the functional whisker representation, defined in this manner, directly overlies the centre of the corresponding layer IV barrel²⁹.

Analysis of electrophysiological recordings

Receptive field and response magnitude measurements were derived from single and multiunit responses to stimulation of 25 whiskers on the contralateral face (AB–A3, BC–B3, CD–C5, DE–D5, E1–E5) independently presented in a pseudorandom order. Whiskers were displaced 1.9° rostrocaudally at 5 Hz for 1 s with a glass capillary tube, attached to a fast piezoelectric biomorph wafer for 40 stimulus repetitions. Receptive field size was calculated by counting the number of whiskers that evoked a response, exceeding a $P < 0.01$ significance threshold (determined by a Poisson probability function applied to spontaneous activity, during the 500 ms preceding stimulus onset) during a 50 ms period, beginning 7 ms after stimulus onset³⁰. Mean spontaneous firing rate was defined as the average firing rate during a 300 ms period, beginning 350 ms before stimulus onset. The magnitude of the stimulus-evoked response was calculated by subtracting the mean spontaneous firing rate from activity occurring in a 50 ms period, beginning 7 ms after stimulus onset. Measurements of evoked response magnitude were only calculated from units that had a receptive field size of at least one whisker. The principal whisker was defined in these units as the whisker that evoked the highest magnitude response. Adjacent whisker response strength was obtained from the average of all adjacent whiskers along the row and arc.

Morphometric analysis of cytochrome oxidase reactivity

Cytochrome oxidase reactivity was measured in coronal and tangential sections. A digital image of each section was obtained at $\times 4$ and $\times 20$ magnification. Cortical thickness was defined as the average distance between the white matter and cortical surface in sections drawn from anterior, central and posterior regions of the barrel cortex. The PMBSF was defined in tangential sections as the region of the barrel cortex containing barrels AB–A3, BC–B3, CD–C5, DE–D5, E1–E5 (Fig. 4a). Three categories of measurements were obtained from tangential sections in NH ($n = 9$) and SC rats ($n = 10$): (1) the average length of the entire PMBSF was measured across arcs (5, 4, 3, 2, 1 and straddler) and rows (A, B, C, D and E); (2) the width, length and area of all 25 barrels; (3) the length of the septa separating all pairs of adjacent barrels. All measurements were performed with measurement tools in Photoshop (Adobe Systems).

Data collection could not be performed blind due to changes in the appearance of the NH rats, but analysis of the cytochrome oxidase maps and electrophysiological data were performed blind to treatment group.

Received 5 December 2003; accepted 2 March 2004; doi:10.1038/nature02469.

- Buonomano, D. V. & Merzenich, M. M. Cortical plasticity: from synapses to maps. *Annu. Rev. Neurosci.* **21**, 149–186 (1998).
- Kaas, J. H. in *The New Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 223–236 (MIT Press, Cambridge, Massachusetts, 2000).
- Polley, D. B., Chen-Bee, C. H. & Frostig, R. D. Two directions of plasticity in the sensory-deprived adult cortex. *Neuron* **24**, 623–637 (1999).
- Simons, D. J. & Land, P. W. Early experience of tactile stimulation influences organization of somatic sensory cortex. *Nature* **326**, 694–697 (1987).
- Fox, K. A critical period for experience-dependent synaptic plasticity in rat barrel cortex. *J. Neurosci.* **12**, 1826–1838 (1992).
- Diamond, M. E., Armstrong-James, M. & Ebner, F. F. Experience-dependent plasticity in adult rat barrel cortex. *Proc. Natl Acad. Sci. USA* **90**, 2082–2086 (1993).
- Glazewski, S. & Fox, K. Time course of experience-dependent synaptic potentiation and depression in barrel cortex of adolescent rats. *J. Neurophysiol.* **75**, 1714–1729 (1996).
- Kossut, M., Hand, P. J., Greenberg, J. & Hand, C. L. Single vibrissal cortical column in SI cortex of rat and its alterations in neonatal and adult vibrissa-deafferented animals: a quantitative 2DG study. *J. Neurophysiol.* **60**, 829–852 (1988).
- Woolsey, T. A. & Van der Loos, H. The structural organization of layer IV in the somatosensory region (SI) of mouse cerebral cortex. The description of a cortical field composed of discrete cytoarchitectonic units. *Brain Res.* **17**, 205–242 (1970).
- Van der Loos, H. & Woolsey, T. A. Somatosensory cortex: structural alterations following early injury to sense organs. *Science* **179**, 395–398 (1973).
- Jeanmonod, D., Rice, F. L. & Van der Loos, H. Mouse somatosensory cortex: alterations in the barrel field following receptor injury at different early postnatal ages. *Neuroscience* **6**, 1503–1535 (1981).
- Bennett, E. L., Diamond, M. C., Krech, D. & Rosenzweig, M. R. Chemical and anatomical plasticity of brain. *Science* **146**, 610–619 (1964).
- Uylings, H. B., Kuypers, K., Diamond, M. C. & Veltman, W. A. Effects of differential environments on plasticity of dendrites of cortical pyramidal neurons in adult rats. *Exp. Neurol.* **62**, 658–677 (1978).
- Green, E. J., Greenough, W. T. & Schlumpf, B. E. Effects of complex or isolated environments on cortical dendrites of middle-aged rats. *Brain Res.* **264**, 233–240 (1983).
- Land, P. W. & Simons, D. J. Cytochrome oxidase staining in the rat Sml barrel cortex. *J. Comp. Neurol.* **238**, 225–235 (1985).
- Fanselow, E. E. & Nicolelis, M. A. Behavioural modulation of tactile responses in the rat somatosensory system. *J. Neurosci.* **19**, 7603–7616 (1999).
- Castro-Alamancos, M. A. Role of thalamocortical sensory suppression during arousal: focusing sensory inputs in neocortex. *J. Neurosci.* **22**, 9651–9655 (2002).
- Castro-Alamancos, M. A. & Oldford, E. Cortical sensory suppression during arousal is due to the activity-dependent depression of thalamocortical synapses. *J. Physiol. (Lond.)* **541**, 319–331 (2002).
- Beaulieu, C. & Cynader, M. Effect of the richness of the environment on neurons in cat visual cortex. I. Receptive field properties. *Brain Res. Dev. Brain Res.* **53**, 71–81 (1990).
- Coq, J. O. & Xerri, C. Environmental enrichment alters organizational features of the forepaw representation in the primary somatosensory cortex of adult rats. *Exp. Brain Res.* **121**, 191–204 (1998).
- Welker, E., Rao, S. B., Dorfl, J., Melzer, P. & van der Loos, H. Plasticity in the barrel cortex of the adult mouse: effects of chronic stimulation upon deoxyglucose uptake in the behaving animal. *J. Neurosci.* **12**, 153–170 (1992).
- Knott, G. W., Quairiaux, C., Genoud, C. & Welker, E. Formation of dendritic spines with GABAergic synapses induced by whisker stimulation in adult mice. *Neuron* **34**, 265–273 (2002).
- Prakash, N., Cohen-Cory, S. & Frostig, R. D. Rapid and opposite effects of BDNF and NGF on the functional organization of the adult cortex *in vivo*. *Nature* **381**, 702–706 (1996).
- Stern, E. A., Maravall, M. & Svoboda, K. Rapid development and plasticity of layer 2/3 maps in rat barrel cortex *in vivo*. *Neuron* **31**, 305–315 (2001).
- Zardecki, P. et al. Synaptic mechanisms of cortical representational plasticity: somatosensory and corticocortical EPSPs in reorganized raccoon SI cortex. *J. Neurophysiol.* **69**, 1422–1432 (1993).
- Fox, K. The cortical component of experience-dependent synaptic plasticity in the rat barrel cortex. *J. Neurosci.* **14**, 7665–7679 (1994).
- Frostig, R. D., Chen-Bee, C. H. & Polley, D. B. in *In Vivo Optical Imaging of Brain Function* (ed. Frostig, R. D.) 21–42 (CRC Press, Boca Raton, 2002).
- Chen-Bee, C. H. et al. Visualizing and quantifying evoked cortical activity assessed with intrinsic signal imaging. *J. Neurosci. Methods* **97**, 157–173 (2000).
- Brett-Green, B., Chen-Bee, C. H. & Frostig, R. D. Comparing the functional representations of central and border whiskers in rat primary somatosensory cortex. *J. Neurosci.* **21**, 9944–9954 (2001).
- Abeles, M. Quantification, smoothing, and confidence limits for single-units' histograms. *J. Neurosci. Methods* **5**, 317–325 (1982).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We would like to thank L. Sininger, J. Rickert and C. Chen-Bee for assistance with data analysis and maintaining the naturalistic habitat, and D. Simmons and P. Yahr for their histology expertise. We would like to thank S. Bao, T. Carew, M. Leon, C. Moore, F. Strata and N. Weinberger for comments on the manuscript and M. Leon for suggestions on the use of a NH. This work was supported by National Institute of Health-National Institute of Neurological Disorders and Stroke grants to R.D.F.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.P. (dpolley@phy.ucsf.edu).

Functional variation in *LGALS2* confers risk of myocardial infarction and regulates lymphotoxin- α secretion *in vitro*

Kouichi Ozaki¹, Katsumi Inoue², Hiroshi Sato³, Aritoshi Iida⁴, Yozo Ohnishi¹, Akihiro Sekine⁴, Hideyuki Sato³, Keita Odashiro², Masakiyo Nobuyoshi², Masatsugu Hori³, Yusuke Nakamura^{4,5} & Toshihiro Tanaka¹

¹Laboratory for Cardiovascular Diseases, SNP Research Center, The Institute of Physical and Chemical Research (RIKEN), Tokyo 108-8639, Japan

²Department of Cardiology, Kokura Memorial Hospital, Kitakyushu 802-8555, Japan

³Department of Internal Medicine and Therapeutics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan

⁴Laboratory for Genotyping, SNP Research Center, The Institute of Physical and Chemical Research (RIKEN), Yokohama 230-0045, Japan

⁵Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan

Myocardial infarction (MI) has become one of the leading causes of death in the world. Its pathogenesis includes chronic formation of plaque inside the vessel wall of the coronary artery and acute rupture of the artery, implicating a number of inflammation-mediating molecules, such as the cytokine lymphotoxin- α (LTA)¹. Functional variations in *LTA* are associated with susceptibility to MI². Here we show that LTA protein binds to galectin-2, a member of the galactose-binding lectin family³. Our case-control association study in a Japanese population showed that a single nucleotide polymorphism in *LGALS2* encoding galectin-2 is significantly associated with susceptibility to MI. This genetic substitution affects the transcriptional level of galectin-2 *in vitro*, potentially leading to altered secretion of LTA, which would then affect the degree of inflammation; however, its relevance to other populations remains to be clarified. Smooth muscle cells and macrophages in the human atherosclerotic lesions expressed both galectin-2 and LTA. Our findings thus suggest a link between the LTA cascade and the pathogenesis of MI.

To understand better the role of LTA in the pathogenesis of this disease, we searched for proteins that interact with LTA. Using the *Escherichia coli* two-hybrid system and the phage display method, we identified galectin-2 as a binding partner of LTA. We purified the two recombinant proteins using a bacterial expression system, and confirmed direct binding of galectin-2 to LTA using an *in vitro* binding assay (Fig. 1a). We further examined their interaction in mammalian cells using constructs designed to express Flag-tagged LTA or S-tagged galectin-2 and western blot analysis; co-immunoprecipitation of galectin-2 and LTA confirmed their interaction (Fig. 1b). Using antibodies specific to each protein, we also investigated subcellular localization of native galectin-2 and LTA in U937 cells, and found that these proteins co-localized in the cytoplasm (Fig. 1c).

We then examined whether any genetic variation in *LGALS2* was associated with susceptibility to MI. Re-sequencing the *LGALS2* genomic region using DNA from 32 MI samples revealed 17 additional single nucleotide polymorphisms (SNPs) (Fig. 2a). Next, we compared the genotype frequencies of approximately 600 individuals with MI, and 600 controls at these SNP loci, and found that one SNP (3279C $\rightarrow$ T) in intron-1 of *LGALS2* was significantly associated with MI (designated SNP9 in Supplementary Table 1). This association was confirmed by increasing the number of samples to 2,302 for patients with MI and 2,038 for

controls, and by including another set of samples (Table 1 and Supplementary Table 2).

We also analysed linkage disequilibrium using a subset of markers with minor allele frequency of >0.20 , and investigated haplotype structure within the *LGALS2* region (Supplementary Table 3). This allowed us to identify a haplotype block containing three SNPs (SNP9, SNP10 and SNP11); no particular haplotype showed a higher statistical significance for association with MI than the single SNP alone (Supplementary Table 4). Because the minor allele frequency of the associated SNP was lower in the patients' group (Table 1), we concluded from these genetic studies that the minor variant protects against the risk of MI, although our study included only Japanese subjects and so its relevance to other populations remains to be clarified.

To investigate the biological significance of this genetic variation in intron-1 of *LGALS2*, we constructed reporter plasmids with a genomic fragment containing the SNP downstream of a luciferase gene and the SV40 enhancer sequence, and examined the effect of the intron-1 SNP on gene expression. The clone containing the 3279T allele showed nearly 50% less transcriptional activity than those containing the 3279C allele or the vector alone (Fig. 2b, c). These observations indicated that the nucleotide substitution could potentially reduce the level of the galectin-2 transcript.

We then hypothesized that the amount of intracellular galectin-2 might regulate the extracellular secretion level of LTA, thereby influencing the degree of inflammation. To test this hypothesis, we examined the effect of the level of galectin-2 on the secreted level of LTA in the medium by repressing galectin-2 expression using a small interference (si)RNA technique, and by over-expressing

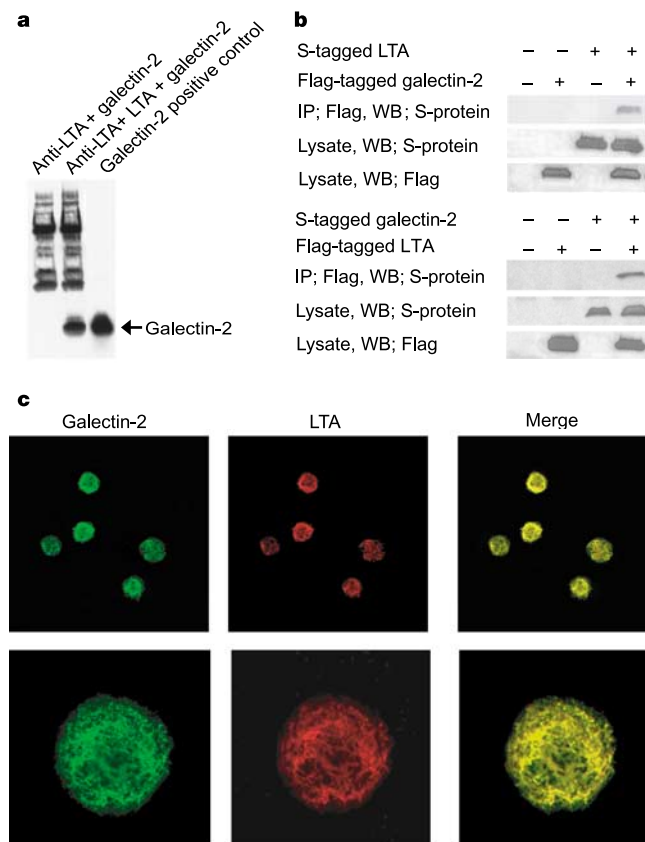


Figure 1 LTA binds to galectin-2. **a**, *In vitro* binding assay. **b**, Co-immunoprecipitation of tagged LTA and galectin-2 in COS7 cells. **c**, Co-localization of endogenous LTA with galectin-2 in U937 cells (top row) with enlarged images of representative cells in the upper panels (bottom row). IP, immunoprecipitation; WB, western blot.

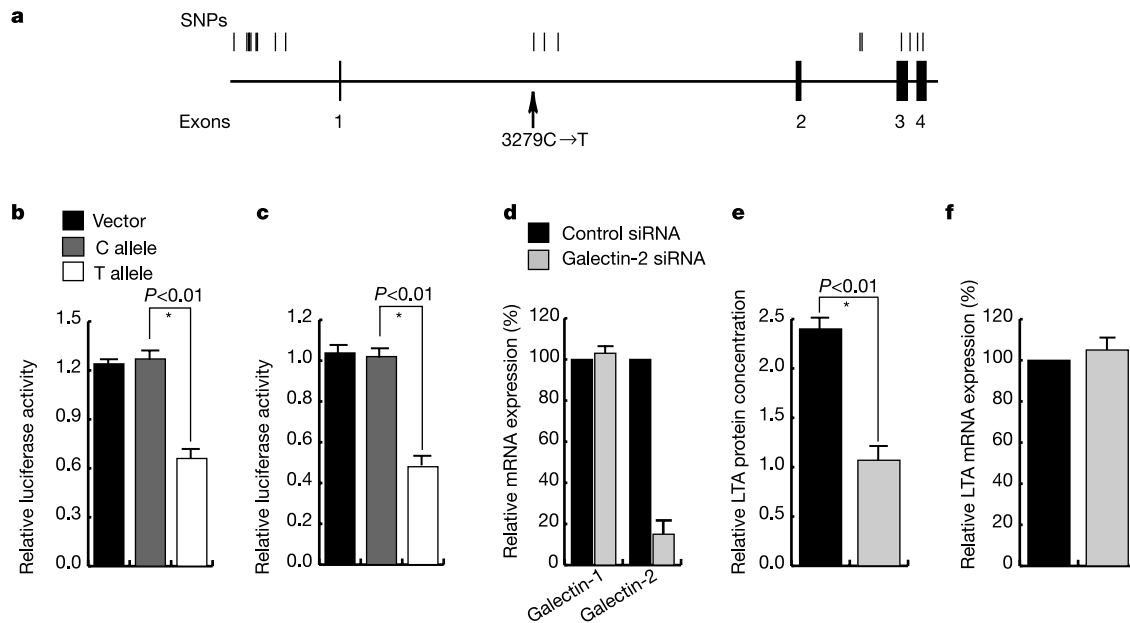


Figure 2 Association of a SNP in *LGALS2* with myocardial infarction and its functionality. **a**, Map of SNPs in *LGALS2* locus. **b**, **c**, Transcriptional regulatory activity of intron-1 SNP of *LGALS2* in HeLa (**b**) and HepG2 (**c**) cells. **d–f**, Inhibition of galectin-2 expression. Levels

of galectin-1 or galectin-2 mRNA (**d**), supernatant LTA protein (**e**), and LTA mRNA (**f**) after 48 h transfection with siRNA vectors. *Student's *t*-test. Each experiment was repeated three times and each sample was studied in triplicate.

galectin-2. One siRNA for galectin-2 repressed galectin-2 messenger RNA to nearly one-fifth of its original level (Fig. 2d) and resulted in inhibition of LTA secretion into the medium (Fig. 2e). Over-expression of galectin-2 enhanced LTA secretion (Supplementary Fig. 1a). As shown in Fig. 2f, the LTA mRNA level was unchanged by knockdown or over-expression of galectin-2 (see also Supplementary Fig. 1b).

To further investigate the regulatory mechanism of LTA secretion by galectin-2, we searched for intracellular molecules that associate with galectin-2, using a tandem affinity purification (TAP) system⁴. We identified two unique bands that could be observed only when the galectin-2-TAP tag was expressed (Fig. 3a). On the basis of a MALDI/TOF (matrix-assisted laser desorption/ionization–time of flight) mass spectrometry analysis, these two bands were shown to correspond to α - and β -tubulins—important components of microtubules. Using HeLa cells transfected with a plasmid to express Flag-tagged galectin-2, we confirmed co-immunoprecipitation of endogenous tubulins and galectin-2 (Fig. 3b and data not shown). Interestingly, the tubulins were also co-immunoprecipitated with LTA (Fig. 3b and data not shown). Images from serial confocal sections of double-immunostained U937 cells revealed

that galectin-2 and α -tubulin co-localized as reticular filamentous networks developed in the cytoplasm (Fig. 3c). Recently, microtubule cytoskeleton networks have been implicated in the subcellular transport of some proteins, including glucose transporter isoform (GLUT4) or thiamine transporter (THTR1)^{5,6}. It is likely that LTA is another molecule that uses the microtubule cytoskeleton network for translocation. It is also conceivable that galectin-2 has a role in intracellular trafficking, although the precise role of galectin-2 in this trafficking machinery complex has yet to be elucidated.

To examine whether these proteins are expressed in the lesion of MI—that is, in atherosclerotic lesions of the coronary artery—and if so, to investigate in which part of the lesion they are expressed, we performed immunohistochemical staining of human coronary atherectomy specimens with anti-LTA or anti-galectin-2 antibody. As shown in Fig. 4a, b, immunoreactivities for both LTA and galectin-2 were detected in intimal cells in atherosclerotic plaques, some of which were spindle-shaped or contained vacuolated, round cytoplasm. Immunostaining of adjacent sections with anti-smooth muscle cell (SMC) α -actin or anti-CD68 showed that the majority of these cells were either SMCs or SMC-derived foam cells, with

Table 1 Association between MI and intron-1 SNP in *LGALS2*

Genotype	Number of patients with MI	Number of patients with MI (%)	Number of controls	Number of controls (%)
<i>LGALS2</i> intron-1 3279C → T				
CC	1,077	46.8	856	42.0
CT	1,014	44.0	903	44.3
TT	211	9.2	279	13.7
Total	2,302	100.0	2,038	100.0
	χ^2	<i>P</i> value	Odds ratio	95% CI
Genotype frequency	25.2	0.0000034		
Allele frequency	21.1	0.0000045	1.23	1.13–1.35
CC versus others	10.0	0.0016	1.21	1.08–1.37
TT versus others	22.1	0.0000026	1.57	1.30–1.90

Nucleotide numbering is according to the mutation nomenclature¹³. CI, confidence interval.

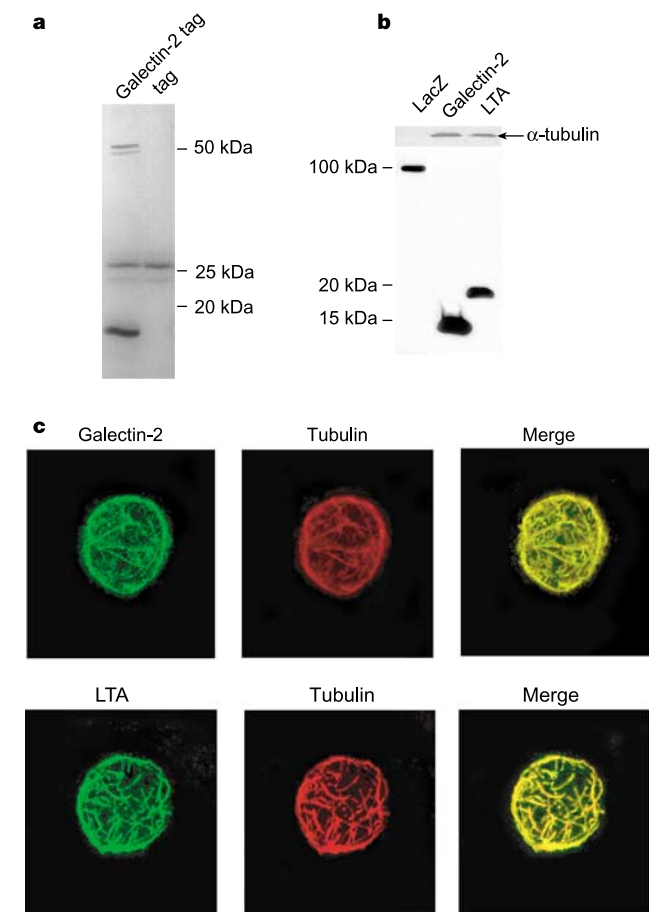


Figure 3 Galectin-2 interacts with microtubules. **a**, Isolation of TAP-tagged galectin-2 and interacting proteins. Arrowheads indicate α - and β -tubulins, revealed by MALDI/TOF mass spectrometry analyses. **b**, Co-immunoprecipitation of endogenous α -tubulin with Flag-tagged galectin-2 or LTA. Immunoprecipitations were done using anti-Flag M2 agarose, and immunoprecipitates were detected using anti- α -tubulin antibody (upper panel) or anti-Flag antibody (lower panel). Flag-tagged LacZ encoding β -galactosidase was used as a negative control. **c**, Co-localization of endogenous α -tubulin with endogenous galectin-2 or LTA in U937 cells.

occasional macrophages being observed (Fig. 4c, d and Supplementary Fig. 2). Co-expression of LTA and galectin-2 was also observed in the majority of polymorphic SMCs by double-labelled immunohistochemistry (Fig. 4e). In contrast, expression of either protein was not detectable in atrophic SMCs of fibrous plaques with scanty cellularity or in normal medial SMCs (Fig. 4f for LTA; data not shown for galectin-2).

These results showed that LTA and galectin-2 were co-expressed in SMCs and macrophages in the intima of human atherosclerotic plaques, but were absent in quiescent or normal medial SMCs. We also confirmed their co-expression by counting immunohistochemically double-labelled cells (Supplementary Table 5). A recent report also indicated that LTA was expressed in atherosclerotic lesions in mice and that loss of LTA reduced the size of the lesions⁷. Together, these results indicate that LTA, as one of the mediators of inflammation, along with galectin-2, as a regulator of LTA secretion, might have roles in the pathogenesis of MI, although the functional correlation of galectin-2 and LTA with the development, progression or rupture of atherosclerotic plaques has yet to be clarified.

We have identified galectin-2 as another risk factor for MI. We believe that the results of this study provide an anchoring point for better understanding of the pathogenesis of MI. □

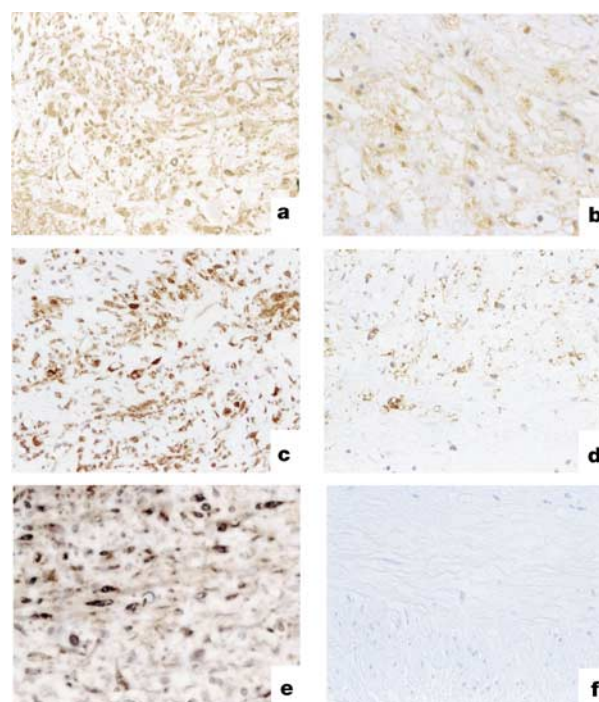


Figure 4 Expression and co-localization of galectin-2 and LTA in the coronary atherectomy specimen. Single-labelled immunohistochemistry of serial sections of primary atherosclerotic lesions from human coronary arteries obtained by directional coronary atherectomy, stained with anti-human LTA (**a**), anti-human galectin-2 (**b**), monoclonal anti-SMC α -actin (**c**), and monoclonal anti-CD68 (**d**). Magnification, $\times 100$. **e**, Double-labelled immunohistochemical staining with anti-LTA (brown) and galectin-2 (blue) antibodies. Magnification, $\times 119$. **f**, Single-labelled immunohistochemical staining with anti-human LTA for atrophic SMCs of fibrous plaques with scanty cellularity and normal medial SMCs. Magnification, $\times 100$.

Methods

DNA samples

The study included 2,638 Japanese patients with MI, referred to as the Osaka Acute Coronary Insufficiency Study (OACIS) group. The diagnosis of definite MI has been described previously². The control subjects consisted of 2,499 general populations recruited through several medical institutes in Japan. All subjects were Japanese and gave written informed consent to participation in the study; or if they were under 20 years old, their parents gave consent according to the process approved by the Ethical Committee at the SNP Research Center, The Institute of Physical and Chemical Research (RIKEN), Yokohama.

SNP analysis

Design for polymerase chain reaction (PCR) primers, PCR experiments, DNA extraction, DNA sequencing, SNP discovery, genotyping of SNPs and statistical analysis has been described previously^{8–10}.

E. coli two-hybrid and phage display screening

We constructed a two-hybrid complementary DNA (cDNA) library using mRNA isolated from Jurkat cells (RIKEN Cell Bank; RCB0806) and BacterioMatch two-hybrid system library construction kit (Clontech), and screened the library using pBT-human LTA as the 'bait'. For the phage display system, we constructed a phage display cDNA library using mRNA isolated from Jurkat cells and T7 select cDNA cloning system (Novagen). We screened the library using the immobilized LTA as the 'bait'.

Tandem affinity purification

The tandem affinity purification procedure was carried out essentially as described in ref. 4, with some modifications. We constructed a fusion cassette encoding a His tag, a TEV cleavage site, and an S tag as a TAP tag sequence in pCMV-Myc vector (Sigma). This TAP vector expresses amino-terminal Myc-tagged target protein with carboxy-terminal TAP tags in mammalian cells under the control of cytomegalovirus promoter. The TAP vector was transiently transfected into HeLa cells (Health Science Research Resources Bank (HSRRB); JCRB9004). Target protein bands were analysed by MALDI/TOF mass spectrometry at APRO Life Science.

In vitro binding assay and co-immunoprecipitation experiment

We prepared purified S-tagged recombinant LTA and T7-tagged galectin-2 derived from *E. coli* using the pET system (Novagen), and combined them. The co-immunoprecipitation experiments were performed using a monoclonal antibody against LTA (R&D Systems) coupled to HiTrap™ NHS-activated Sepharose HP (Amersham). We visualized the immune complex using T7 tag antibody (Stratagene) and horseradish peroxidase (HRP) conjugated with anti-mouse IgG antibody. For co-immunoprecipitation in mammalian cells, we transfected expression plasmids of Flag or S-tagged LTA, galectin-2 and LacZ (as a negative control) into COS7 cells (HSRRB; JCRB9127) or HeLa cells using Fugene. Immunoprecipitations were done in lysis buffer (20 mM Tris pH 7.5, with 150 mM NaCl, 0.1 % Nonident P-40). Twenty-four hours after transfection, cells were lysed, and immunoprecipitations were performed using anti-Flag tag M2 agarose (Sigma). We visualized the immune complex using HRP-conjugated S-protein (Novagen), anti-Flag M2 peroxidase conjugate (Sigma) or mouse monoclonal antibody against human α -tubulin (Molecular Probes) and HRP-conjugated anti-mouse IgG antibody.

Confocal microscopy

Polyclonal anti-human galectin-2 antisera were raised in rabbits using recombinant protein synthesized in *E. coli*. The antisera showed no cross-reactivity to structurally related molecules galectin-1 and galectin-3, analysed by western blot. Polyclonal anti-galectin-2 antisera and either goat anti-human LTA IgG (R&D Systems) or mouse anti-human α -tubulin monoclonal IgM antibodies were used with Alexa secondary antibodies (Molecular Probes). U937 cells (HSRRB; JCRB9021) were stimulated for 30 min with phorbol myristate acetate (PMA) (20 ng ml^{-1}) and fixed. They were subsequently incubated with the corresponding primary antibodies in phosphate-buffered saline containing 3% bovine serum albumin, and the corresponding Alexa secondary antibodies.

siRNA and over-expression experiments

The target sequences for galectin-2 (5'-AATCCACCATTGTCTGCAACT-3') were cloned into pSilencer 2.0-U6 siRNA vector (Ambion). For the over-expression experiment, the galectin-2 was cloned into pFlag-CMV5a vector. After transfection, Jurkat cells were stimulated with PMA (20 ng ml^{-1}) for 24 h, and cells and supernatants were collected separately. LTA concentration was measured using an LTA-specific ELISA system (R&D Systems), and normalized by comparison with total protein concentration. The mRNA quantification procedure has been described previously².

Luciferase assay

A DNA fragment, corresponding to nucleotides 3,188 to 3,404 of intron-1 of *LGALS2*, was cloned into pGL3-enhancer vector (Promega) in the downstream of SV40 enhancer in the 5' to 3' orientation. After 24 h transfection, luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega).

Immunohistochemistry

Tissue samples were obtained from 16 patients with MI by elective directional coronary atherectomy. Immunohistochemical protocols were carried out as described previously^{1,12} using goat anti-human LTA IgG (R&D Systems) and rabbit polyclonal anti-human galectin-2 antibody. Staining of adjacent sections was carried out using human-cell-type-specific monoclonal antibodies against SMC 2-actin and CD68 (DAKO). For double-labelled immunohistochemistry, sections were incubated with anti-LTA antibody, then with biotinylated swine anti-goat IgG, and then with avidin-biotin-peroxidase conjugate, followed by visualization with 3,3'-diaminobenzidine tetrahydrochloride (Vector Labs). The section was subsequently incubated with rabbit polyclonal anti-human galectin-2 antibody, followed by incubation with alkaline phosphatase-conjugated swine anti-rabbit IgG and visualized with the 5-bromo-4-chloro-3-indoxyl phosphate and nitro-blue tetrazolium chloride (BCIP/NBT) substrate system.

Received 5 January; accepted 18 March 2004; doi:10.1038/nature02502.

- Ross, R. Atherosclerosis—an inflammatory disease. *N. Engl. J. Med.* **340**, 115–126 (1999).
- Ozaki, K. et al. Functional SNPs in the lymphotoxin- α gene that are associated with susceptibility to myocardial infarction. *Nature Genet.* **32**, 650–654 (2002).
- Gitt, M. A., Massa, S. M., Leffler, H. & Barondes, S. H. Isolation and expression of a gene encoding L-14-II, a new human soluble lactose-binding lectin. *J. Biol. Chem.* **267**, 10601–10606 (1992).
- Rigaut, G. et al. A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
- Liu, L. B., Omatu, W., Kojima, I. & Shibata, H. Insulin recruits GLUT4 from distinct compartments via distinct traffic pathways with differential microtubule dependence in rat adipocytes. *J. Biol. Chem.* **278**, 30157–30169 (2003).
- Subramanian, V. S., Marchant, J. S., Parker, I. & Said, H. M. Cell biology of the human thiamine transporter-1 (hTHT1). Intracellular trafficking and membrane targeting mechanisms. *J. Biol. Chem.* **278**, 3976–3984 (2003).
- Schreyer, S. A., Vick, C. M. & LeBoeuf, R. C. L. Loss of lymphotoxin- α but not tumor necrosis factor- α reduces atherosclerosis in mice. *J. Biol. Chem.* **277**, 12364–12368 (2002).
- Iida, A. et al. Catalog of 258 single-nucleotide polymorphisms (SNPs) in genes encoding three organic anion transporters, three organic anion-transporting polypeptides, and three NADH:ubiquinone oxidoreductase flavoproteins. *J. Hum. Genet.* **46**, 668–683 (2001).
- Ohnishi, Y. et al. A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* **46**, 471–477 (2001).
- Yamada, R. et al. Association between a single-nucleotide polymorphism in the promoter of the human interleukin-3 gene and rheumatoid arthritis in Japanese patients, and maximum-likelihood estimation of combinatorial effect that two genetic loci have on susceptibility to the disease. *Am. J. Hum. Genet.* **68**, 674–685 (2001).

- Minami, M. et al. Expression of SR-PSOX, a novel cell-surface scavenger receptor for phosphatidylserine and oxidized LDL in human atherosclerotic lesions. *Arterioscler. Thromb. Biol.* **21**, 1796–1800 (2001).
- Shi, S. R., Key, M. E. & Kalra, K. L. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *J. Histochem. Cytochem.* **39**, 741–748 (1991).
- den Dunnen, J. T. & Antonarakis, S. E. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum. Mutat.* **15**, 7–12 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Takahashi, M. Yoshii, M. Omotezako, Y. Arijii, S. Abiko, W. Yamanobe and K. Tabei for their assistance. We also thank all the other members of the SNP Research Center, RIKEN, for their contribution to the completion of our study. This work was supported by a grant from the Japanese Millennium Project.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.T. (toshitan@ims.u-tokyo.ac.jp).

Differential modulation of endotoxin responsiveness by human caspase-12 polymorphisms

Maya Saleh¹, John P. Vaillancourt¹, Rona K. Graham², Matthew Huyck³, Srinivasa M. Srinivasula⁴, Emad S. Alnemri⁴, Martin H. Steinberg³, Vikki Nolan³, Clinton T. Baldwin³, Richard S. Hotchkiss⁵, Timothy G. Buchman⁵, Barbara A. Zehnauer⁶, Michael R. Hayden², Lindsay A. Farrer³, Sophie Roy¹ & Donald W. Nicholson¹

¹Department of Biochemistry, Molecular Biology and Pharmacology, Merck Frosst Centre for Therapeutic Research, Montreal, Quebec H9H 3L1, Canada

²Center for Molecular Medicine and Therapeutics, Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V5Z 4H4, Canada

³Departments of Medicine (Genetics Program & Hematology/Oncology Section), Neurology and Genetics & Genomics, and Center for Human Genetics, Boston University School of Medicine and Departments of Epidemiology and Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA

⁴Department of Microbiology and Immunology, Kimmel Cancer Institute, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA

⁵Department of Anesthesiology, Department of Surgery, and ⁶Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA

Caspases mediate essential key proteolytic events in inflammatory cascades and the apoptotic cell death pathway. Human caspases functionally segregate into two distinct subfamilies: those involved in cytokine maturation (caspase-1, -4 and -5) and those involved in cellular apoptosis (caspase-2, -3, -6, -7, -8, -9 and -10)^{1,2}. Although caspase-12 is phylogenetically related to the cytokine maturation caspases, in mice it has been proposed as a mediator of apoptosis induced by endoplasmic reticulum stress including amyloid- β cytotoxicity, suggesting that it might contribute to the pathogenesis of Alzheimer's disease³. Here we show that a single nucleotide polymorphism in caspase-12 in humans results in the synthesis of either a truncated protein (Csp12-S) or a full-length caspase proenzyme (Csp12-L). The read-through single nucleotide polymorphism encoding Csp12-L is confined to populations of African descent and confers hypo-responsiveness to lipopolysaccharide-stimulated cytokine production in *ex vivo* whole blood, but has no significant effect on apoptotic sensitivity. In a preliminary study, we find that the frequency of the Csp12-L

allele is increased in African American individuals with severe sepsis. Thus, Csp12-L attenuates the inflammatory and innate immune response to endotoxins and in doing so may constitute a risk factor for developing sepsis.

While cloning human caspase-12 complementary DNAs and sequencing their corresponding genomic DNA, we found that almost all clones contained a TGA (stop) codon at amino acid position 125, as previously described⁴, but a few DNA sources contained a read-through CGA (Arg) codon instead (Fig. 1). This single nucleotide polymorphism (SNP) was contained in exon 4 of the gene encoding human caspase-12 (which was found to be proximal to the caspase-1, -4, -5 gene cluster on 11q23) and resulted in messenger RNAs encoding either a full-length, tripartite caspase

precursor protein (Csp12-L) or a truncated polypeptide (Csp12-S) ending at the junction between the prodomain and the large subunit (Fig. 1a, b). Sequence analysis of more than 1,100 genomic DNA samples from people of distinct ethnic backgrounds showed that most encoded the truncated prodomain-only form of caspase-12 (Csp12-S). The less-frequent CGA (Arg) polymorphism resulting in a full-length caspase polypeptide (Csp12-L) was found only in populations of African descent and was absent in all Caucasian and Asian groups tested (Fig. 1c, d). Although less frequent in humans and confined to about 20% of people of African descent (Fig. 1d and Supplementary Table 1), the full-length form of caspase-12 was found to be encoded in all other species tested, including new world and old world primates and rodents (Supplementary Table 2).

Caspase-12 is naturally polymorphic in ethnic groups of African descent, providing an ideal system in which to examine its *in vivo* role in humans and to circumvent the pitfalls associated with studying caspases in recombinant systems. We chose to examine the apoptotic and inflammatory responsiveness of cells in whole blood obtained from consenting donors of African origin. Owing to the genomic and structural association of caspase-12 with the pro-inflammatory caspase-1, -4, -5 gene cluster, we first examined the effect of the caspase-12 polymorphism on lipopolysaccharide (LPS) and concanavalin A (conA)-stimulated cytokine production in whole blood (Fig. 2). *In vivo* expression of Csp12-L in blood cells was confirmed by western blotting, and the protein was induced by LPS treatment (Fig. 2a). For most cytokines examined, the presence of Csp12-L reduced the magnitude of the LPS-induced response: maximum attenuation occurred in people who were homozygous for the T125C allele (Csp12-L/L) and an intermediate response occurred in heterozygotes (Csp12-S/L) as compared with T125 (Csp12-S/S) homozygotes (Fig. 2b, c).

Cytokine production stimulated by conA was unaffected by the two variants of caspase-12 (data not shown), with the exception of interferon- γ , which was substantially increased in response to conA by the presence of one Csp12-L allele and further increased in Csp12-L/L homozygotes as compared with Csp12-S/S controls (Fig. 2d). These results support a role for caspase-12 as a master attenuator of the macrophage-elicited T-helper cell type 1 and type 2 cytokine response, with probable compensatory enabling of T-cell-derived interferon- γ formation. By contrast, the naturally occurring variants of human caspase-12 had no significant effect on apoptotic sensitivity to diverse stimuli, including activators of the extrinsic and intrinsic cell death pathways, as well as agents that provoke apoptosis through endoplasmic reticulum (ER) stress (Fig. 3). These latter findings are contrary to those observed in rodents, where caspase-12 has been proposed to be a key mediator of ER-stress-induced cell death and has been implicated in neurodegenerative disorders including Alzheimer's disease^{3,5}, polyglutamine repeat disorders⁶ and ischaemic brain injury^{7,8}.

These findings suggest that human caspase-12 has a role in modulating endotoxin responsiveness and cytokine release. Because other caspases of this subfamily promote cytokine formation through precursor maturation (for example, caspase-1-mediated cleavage of interleukin-1 β (IL-1 β) and IL-18), an attenuating role of caspase-12 seems counterintuitive. It suggests that full-length caspase-12 (Csp12-L) might act as a dominant-negative regulator of inflammatory caspase activation, potentially by antagonizing the inflammasome complex and associated pro-inflammatory pathways, such as NF- κ B (refs 9–13). In support of this, we found that human caspase-12 was devoid of detectable catalytic activity, in contrast to rodent caspase-12 proteins, which underwent auto-catalytic maturation (data not shown). Furthermore, transfected cell lines expressing Csp12-L showed dampened NF- κ B activation in response to tumour-necrosis factor- α (TNF- α) and reduced

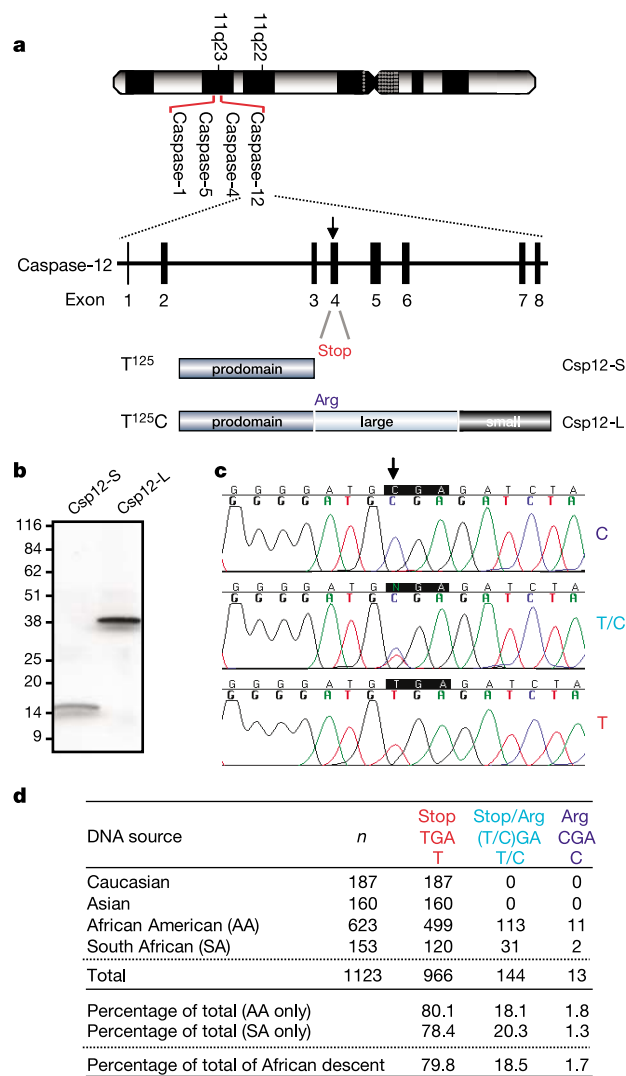


Figure 1 Identification of a caspase-12 point mutation in individuals of African descent. **a**, Map of the region at 11q23. The gene encoding caspase-12 is clustered with those encoding caspase-1, -4 and -5. The exon-intron organization of caspase-12 is shown. Arrow indicates the polymorphism TGA $\rightarrow$ CGA. In wild-type (WT) individuals, a stop codon in exon 4 encodes a prodomain-only protein (Csp12-S). In T125C individuals, an arginine replaces the stop and encodes a full-length protein (Csp12-L). **b**, Western blot. Caspase-12 variants were *in vitro* transcribed and translated, and detected by antibodies specific for the human caspase-12 prodomain. **c**, Electropherograms from control, heterozygous and T125C homozygous individuals. Arrow indicates the position of the polymorphism. **d**, Genotype frequency of T125 and T125C in different ethnic backgrounds.

IL-1-stimulated release of IL-8, an NF- κ B-dependent process (Fig. 4). Having only the CARD domain, Csp12-S was a weaker inhibitor of NF- κ B activation. Collectively, these data indicate that human caspase-12 can function as a dominant-negative regulator of inflammatory responses and innate immunity.

Because human caspase-12 modulated endotoxin responsiveness in *ex vivo* human whole blood but had no effect on apoptotic sensitivity, we carried out studies to examine whether there is an association between the caspase-12 polymorphism and either sepsis or Alzheimer's disease in African Americans. We chose sepsis

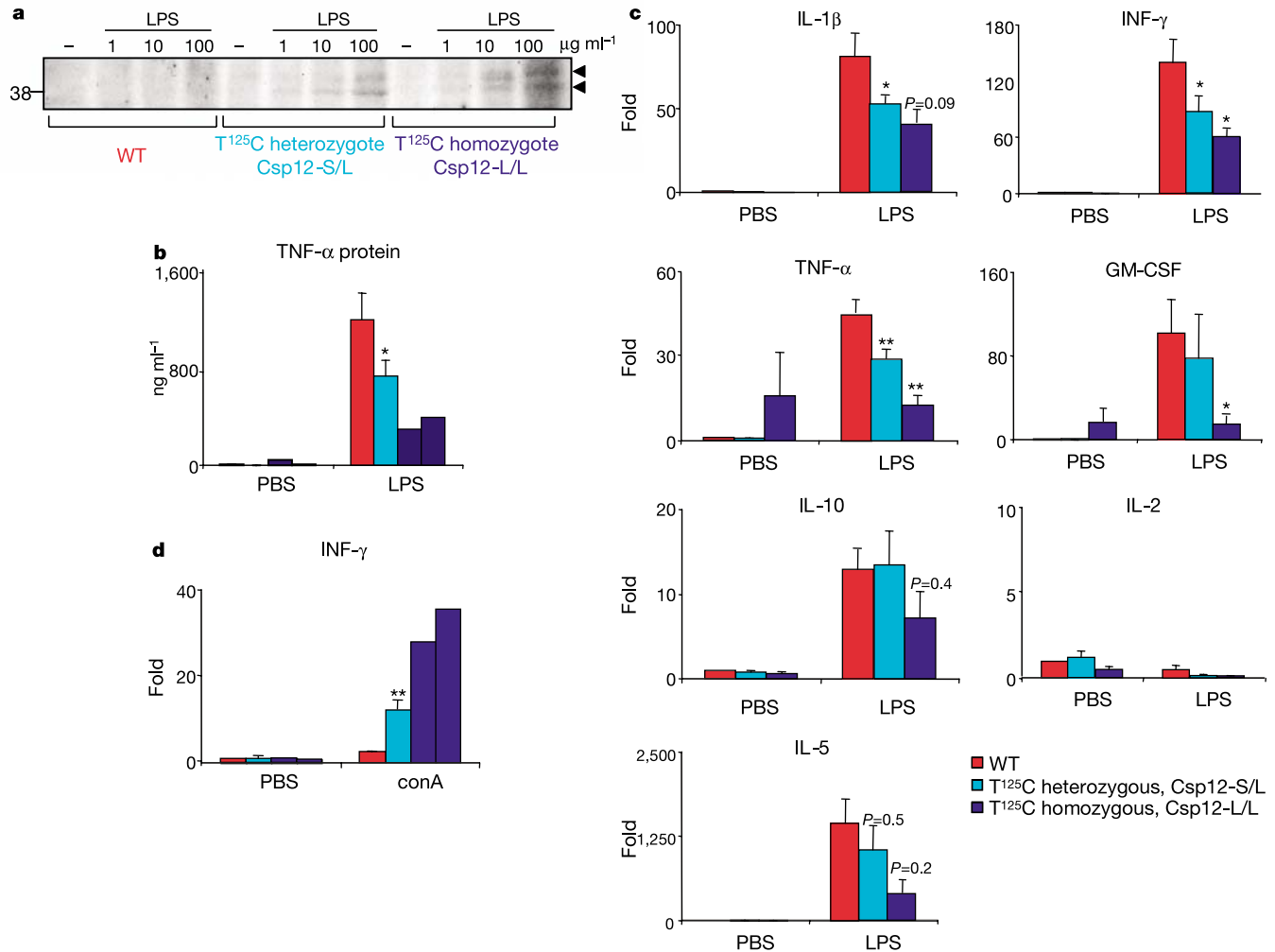


Figure 2 Differential responsiveness to LPS and conA of *ex vivo* whole blood from wild-type, T125C heterozygous and T125C homozygous individuals of African descent. **a**, Csp12-L is expressed in the blood of T125C but not wild-type (WT) individuals. **b–d**, Differential responsiveness to LPS and conA. Whole blood was treated with 1 μ g ml⁻¹ LPS (**b, c**) or 50 μ g ml⁻¹ conA (**d**) for 4 h at 37 °C; serum was then collected for TNF- α ELISA (**b**) or total RNA extracted from white blood cells was used for real-time

PCR quantification of cytokines (**c, d**). GM-CSF, granulocyte-macrophage colony-stimulating factor; IFN- γ , interferon- γ . Red, light blue and dark blue represent data from the blood of wild-type, T125C heterozygous and T125C homozygous individuals, respectively. Values represent the mean $\pm$ s.e.m. (* $P < 0.05$; ** $P < 0.01$). Wild type: $n = 8$ (**b**), $n = 20$ (**c**), $n = 16$ (**d**); T125C heterozygous: $n = 8$ (**b**), $n = 20$ (**c**), $n = 16$ (**d**); T125C homozygous: $n = 2$ (**b**), $n = 4$ (**c**), $n = 2$ (**d**).

Table 1 Caspase-12 genotype and allele frequency in Alzheimer's or severe sepsis

African American DNA source	Stop TGA (n)	Stop/Arg (T/C)GA (n)	Arg CGA (n)	Total (n)	Genotype frequency (%)			Allele frequency (%)	
	T/T	T/C	C/C		T/T	T/C	C/C	T	C
Reference group	499	113	11	623	80.1	18.1	1.8	89.2	10.8
Alzheimer's disease group	141	40	3	184	76.7	21.7	1.6	87.5	12.5
Sibling controls	80	21	2	103	77.7	20.4	1.9	87.9	12.1
Unrelated controls	56	14	1	71	78.9	19.7	1.4	88.7	11.3
Sepsis group	23	11	4	38	60.5	29	10.5	75	25
Unrelated controls	120	26	2	148	81.1	17.6	1.3	89.9	10.1

The reference group includes all African Americans without disease assessed in this study (from Fig. 1d). Controls for Alzheimer's disease were cognitively normal siblings and unrelated volunteers. Data for the controls for severe sepsis were obtained from intensive-care-unit patients in the same study who were not septic plus other unaffected African Americans from the same region. The Fisher exact test was used for statistical analysis.

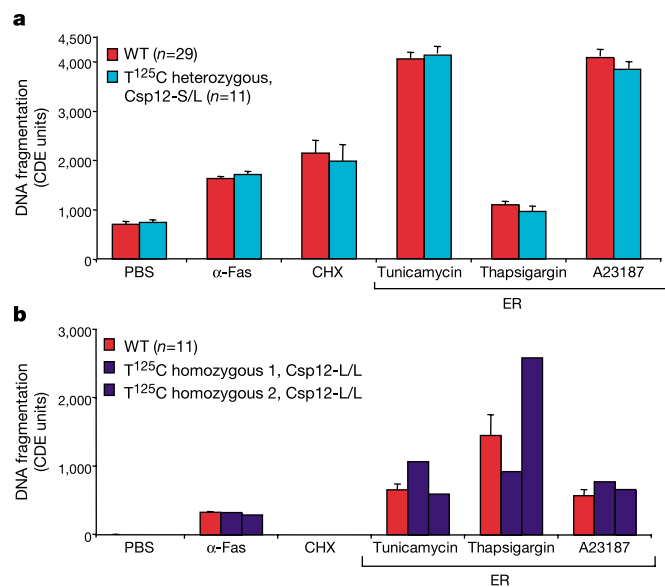


Figure 3 Human caspase-12 is not involved in ER-stress-mediated apoptosis. White blood cells from wild-type (Csp12-S/S), T125C heterozygous (Csp12-S/L) and T125C homozygous (Csp12-L/L) individuals of African descent were treated with the indicated apoptotic stimuli, including three putative ER stressors (ER). After 24 h at 37 °C, cell death by apoptosis was measured by a cell death ELISA. Values represent the mean $\pm$ s.e.m. **a**, An experiment in which no homozygous Csp12-L/L individuals were found in the donor group; **b**, an additional experiment in which two homozygous Csp12-L/L individuals (shown separately) were found.

because of the clear link between this disorder and both perturbed cytokine responsiveness and caspases^{14,15}, and Alzheimer's disease because of the reported resistance of cortical neurons derived from caspase-12-deficient mice to amyloid- β cytotoxicity³. In these preliminary studies, the frequencies of the caspase-12 genotypes and alleles in individuals with Alzheimer's disease were indistinguishable from those of non-affected siblings or unrelated African American age-matched controls (Table 1), consistent with our data from whole blood showing that caspase-12 function in humans is not associated with ER stress and is different to that reported in mice. There was, however, a modest but statistically significant increase in the frequency of genotypes encoding Csp12-L in individuals of African descent who were diagnosed with severe clinical sepsis ($P = 0.005$), including a 7.8-fold increase in Csp12-L/L (T125C/T125C) homozygotes. Occurrence of the T125C allele was roughly doubled in individuals of African descent with sepsis, as compared with all other groups (for example, 25% versus 10.1% for control subjects in the same study; $P = 0.002$). Among individuals of African descent with severe sepsis, the mortality rate was 54% in individuals with a T125C allele as compared with 17% in individuals with only T125 (data not shown). Collectively, these results indicate that the presence of the T125C allele (encoding Csp12-L) may increase susceptibility to severe sepsis and also may result in higher mortality rates (up to threefold) once severe sepsis develops.

In summary, caspase-12 seems to modulate inflammation and innate immunity in humans. More specifically, the full-length caspase-12 polymorph (Csp12-L) confers endotoxin hypo-responsiveness, which seems to be manifest in the clinic as an increased susceptibility to severe sepsis and mortality. Mechanistically, Csp12-L functions as a dominant-negative regulator of essential cellular responses, including the IL-1 and NF- κ B pathways. These findings indicate that caspase-12 antagonists may have therapeutic use in sepsis and other inflammatory and immune disorders,

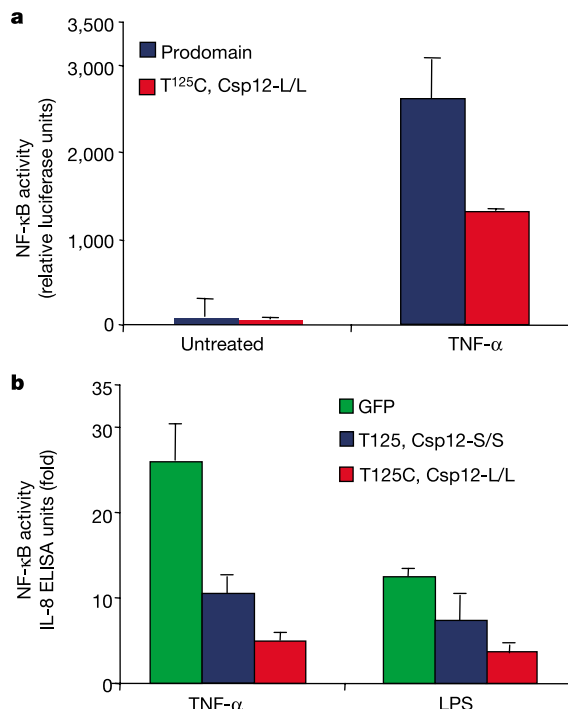


Figure 4 Inhibition of NF- κ B activity by human caspase-12. **a**, HEK 293T cells were co-transfected with the pRSV- β -gal and p κ B-luc reporter plasmids and pcDNA3.1-caspase12(T125) or pcDNA3.1-caspase12(T125C). **b**, HUVEC cells were co-transfected with pEGFP-N1 and pcDNA3.1, pcDNA3.1-caspase12(T125) or pcDNA3.1-caspase12(T125C). Secretion of IL-8 into the culture medium was measured by ELISA. Values represent the mean $\pm$ s.e.m.

where perturbed cytokine responsiveness contributes to disease pathogenesis. □

Methods

Sequencing of caspase-12

Whole blood was collected from humans, squirrel monkeys, capuchins, cynomolgus, rhesus monkeys and Japanese monkeys, and hair was collected from gorillas and chimpanzees. Genomic DNA was extracted from the blood and hair follicles using the QIAamp DNA blood mini kit (Qiagen). We used primers framing a region of 300 base pairs surrounding the T125C polymorphism (sense, 5'-GTCATTCTGTGTATTAA TTGC-3'; antisense, 5'-CCTATAATATCATACATCTTGCTC-3') to amplify the genomic DNA by polymerase chain reaction (PCR). The PCR product was sequenced directly using BigDye Terminators v3.0 (Applied Biosystems).

Blood collection

We collected blood samples from people of African descent through different Black community centres in the Montreal area. For each of five independent experiments, blood donor clinics of roughly 50 donors were organized. We collected 25-ml blood (3 $\times$ 8 ml Vacutainer tubes with heparin; Becton Dickinson) from each donor by venous puncture and pooled the blood from each donor before the start of the *ex vivo* treatments. The people of African descent in this study were of different geographical origin: African, Caribbean, African American and South African. For each individual, informed consent for a molecular genetic study was obtained. The blood samples were collected anonymously. In some experiments, blood cells (red blood cells, total white blood cells, neutrophils, basophils, eosinophils, monocytes, lymphocytes) were counted and found to be unaltered among genotypes.

Ex vivo blood treatment

We treated the blood from all donors first and genotyped subsequently. For the inflammation experiments, whole blood (25 ml) was treated with PBS only (12 ml), 1 μ g ml⁻¹ LPS (*Escherichia coli* 0111:B4, Sigma; 6 ml) or 50 μ g ml⁻¹ conA (*Canavalia ensiformis*, Sigma; 6 ml) for 4 h at 37 °C. After incubation, red blood cells were lysed using erythrocyte lysis buffer (Qiagen) and total RNA was extracted from white blood cells using TRIzol reagent (Gibco-BRL). RNA was used for quantitative real-time PCR of cytokine transcripts. For the cell death experiments, blood was collected in Vacutainer CPT tubes (Becton Dickinson). Mononuclear cells were separated from whole blood by centrifugation and were stimulated with PBS only, 1 μ g ml⁻¹ α -Fas, 1 mM cyclohexamide, 1 μ g ml⁻¹ tunicamycin, 2 μ M thapsigargin or 2 μ M A23187 for 18 h at 37 °C. Cell death was

measured by quantification of oligonucleosomal DNA fragmentation by using Roche's Cell Death enzyme-linked immunoabsorbent assay (ELISA). In all the blood experiments, 200 µl of blood was used for genomic DNA extraction and genotyping (see above).

Real-time quantitative PCR and TNF-α ELISA

Total RNA was prepared by an RNeasy mini kit (Qiagen). Reverse transcription of RNA (50 ng) was done with Taqman transcription reagents (PE Biosystems). We purchased the PCR primers and Taqman probes (PE Biosystems) for the target genes and the 18S ribosomal RNA as pre-developed primers and probe sets (see also Supplementary Information). Plasma TNF-α was quantified by ELISA (Abraxis).

Western blotting

Csp12-L and Csp12-S were *in vitro* transcribed and translated using TNT-coupled reticulocyte lysates (Promega) and were processed for western analysis using rabbit polyclonal antibodies directed against recombinant human caspase-12 prodomain. Alternatively, to detect Csp12-L in blood, isolated white blood cells were lysed in 1 × SDS-PAGE sample buffer and the protein extracts processed for western analysis using rabbit polyclonal antibodies directed against the large subunit of recombinant rat caspase-12.

NF-κB activation assays

For the luciferase assays, we co-transfected HEK 293T cells with κB-luc and β-gal reporter plasmids and a plasmid encoding either Csp12-S (residues 1–125 of the prodomain fused to cMyc) or Csp12-L (T125C caspase-12 fused to green fluorescent protein (GFP)). Twenty-four hours after transfection, cells were treated with 10 ng ml⁻¹ TNF-α for 6 h or were left untreated. Cell extracts were prepared and relative luciferase activity was measured. For the measurement of IL-8 secretion, we transfected HUVEC cells as above, except that the short caspase-12 construct was replaced by the long caspase-12 construct, in which the arginine was mutated to a stop codon at position 125, and pEGFP-N1 was co-transfected as a transfection marker. Twenty-four hours after transfection, the cells were trypsinized and GFP-positive cells were sorted by FACS and allowed to adhere before being treated with PBS, 1 µg ml⁻¹ LPS or 10 ng ml⁻¹ TNF-α for 18 h. The media from the cultured cells was collected and IL-8 was quantified by ELISA.

Human subjects

Individuals with Alzheimer's disease and matched controls were African American participants of the MIRAGE Study, a multicentre family study of genetic and environmental risk factors for Alzheimer's disease¹⁶. All affected individuals met NINCDS/ADRDA criteria¹⁷ for probable or definite Alzheimer's disease. Controls were cognitively normal siblings and unrelated volunteers (including spouses and age-matched members from the same community as the affected individuals). African American individuals with severe sepsis had both septic bacteraemia accompanied by physiological failure of at least one organ system; matched controls were contributors to the Genetic Predisposition to Severe Sepsis (GenPSS) study of the Project IMPACT. We carried out SNP analysis by TDI-FP¹⁸ and sequencing.

Received 2 February; accepted 1 March 2004; doi:10.1038/nature02451.

- Nicholson, D. W. Caspase structure, proteolytic substrates, and function during apoptotic cell death. *Cell Death Differ.* **6**, 1028–1042 (1999).
- Lamkanfi, M., Declercq, W., Kalai, M., Saelens, X. & Vandenabeele, P. Alice in caspase land. A phylogenetic analysis of caspases from worm to man. *Cell Death Differ.* **9**, 358–361 (2002).
- Nakagawa, T. *et al.* Caspase-12 mediates endoplasmic-reticulum-specific apoptosis and cytotoxicity by amyloid-β. *Nature* **403**, 98–103 (2000).
- Fischer, H., Koenig, U., Eckhart, L. & Tschachler, E. Human caspase 12 has acquired deleterious mutations. *Biochem. Biophys. Res. Commun.* **293**, 722–726 (2002).
- Chan, S. L., Culmsee, C., Haughey, N., Klapper, W. & Mattson, M. P. Presenilin-1 mutations sensitize neurons to DNA damage-induced death by a mechanism involving perturbed calcium homeostasis and activation of calpains and caspase-12. *Neurobiol. Dis.* **11**, 2–19 (2002).
- Kourouk, Y. *et al.* Polyglutamine aggregates stimulate ER stress signals and caspase-12 activation. *Hum. Mol. Genet.* **11**, 1505–1515 (2002).
- Shibata, M. *et al.* Activation of caspase-12 by endoplasmic reticulum stress induced by transient middle cerebral artery occlusion in mice. *Neuroscience* **118**, 491–499 (2003).
- Mouw, G. *et al.* Activation of caspase-12, an endoplasmic reticulum resident caspase, after permanent focal ischemia in rat. *NeuroReport* **14**, 183–186 (2003).
- Stehlik, C. *et al.* The PAAD/PYRIN-only protein POP1/ASC2 is a modulator of ASC-mediated nuclear-factor-κB and pro-caspase-1 regulation. *Biochem. J.* **373**, 101–113 (2003).
- Martinon, F., Burns, K. & Tschopp, J. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-β. *Mol. Cell* **10**, 417–426 (2002).
- Srinivasula, S. M. *et al.* The PYRIN-CARD protein ASC is an activating adaptor for caspase-1. *J. Biol. Chem.* **277**, 21119–21122 (2002).
- Grenier, J. M. *et al.* Functional screening of five PYPAF family members identifies PYPAF5 as a novel regulator of NF-κB and caspase-1. *FEBS Lett.* **530**, 73–78 (2002).
- Bouchier-Hayes, L. & Martin, S. J. CARD games in apoptosis and immunity. *EMBO Rep.* **3**, 616–621 (2002).
- Hotchkiss, R. S. *et al.* Caspase inhibitors improve survival in sepsis: a critical role of the lymphocyte. *Nature Immunol.* **1**, 496–501 (2000).
- Hotchkiss, R. S. *et al.* Sepsis-induced apoptosis causes progressive profound depletion of B and CD4⁺ T lymphocytes in humans. *J. Immunol.* **166**, 6952–6963 (2001).
- Green, R. C. *et al.* Risk of dementia among white and African American relatives of patients with Alzheimer disease. *J. Am. Med. Assoc.* **287**, 329–336 (2002).
- McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology* **34**, 939–944 (1984).

- Freeman, B. D., Buchman, T. G., McGrath, S., Tabrizi, A. R. & Zehnauer, B. A. Template-directed dye-terminator incorporation with fluorescence polarization detection for analysis of single nucleotide polymorphisms implicated in sepsis. *J. Mol. Diagn.* **4**, 209–215 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Menard, B. Simpson and the Granby Zoo for non-invasive samples for primate sequencing, and the West Island and Côte Des Neiges Black Community Associations for coordinating blood donor clinics. M.S. is supported by a CIHR postdoctoral fellowship; T.G.B. is supported by a grant from the NIGMS; L.A.F. is supported in part by grants from the NIH.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to D.W.N. (donald_nicholson@merck.com).

Nitration of a peptide phytotoxin by bacterial nitric oxide synthase

Johan A. Kers¹, Michael J. Wach¹, Stuart B. Krasnoff¹, Joanne Widom³, Kimberly D. Cameron¹, Raghida A. Bukhalid^{1*}, Donna M. Gibson², Brian R. Crane³ & Rosemary Loria¹

¹Department of Plant Pathology, Cornell University, Ithaca, New York 14853, USA

²Agricultural Research Service, United States Department of Agriculture, Ithaca, New York 14853, USA

³Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA

* Present address: Bauer Center for Genomics Research, Harvard University, Cambridge, Massachusetts 02138, USA

Nitric oxide (NO) is a potent intercellular signal in mammals that mediates key aspects of blood pressure, hormone release, nerve transmission and the immune response of higher organisms^{1–4}. Proteins homologous to full-length mammalian nitric oxide synthases (NOSs) are found in lower multicellular organisms⁵. Recently, genome sequencing has shown that some bacteria contain genes coding for truncated NOS proteins; this is consistent with reports of NOS-like activities in bacterial extracts^{6,7}. Biological functions for bacterial NOSs are unknown, but have been presumed to be analogous to their role in mammals. Here we describe a gene in the plant pathogen *Streptomyces turgidiscabies* that encodes a NOS homologue, and we reveal its role in nitrating a dipeptide phytotoxin required for plant pathogenicity⁸. High similarity between bacterial NOSs indicates a general function in biosynthetic nitration; thus, bacterial NOSs constitute a new class of enzymes^{9–11}. Here we show that the primary function of *Streptomyces* NOS is radically different from that of mammalian NOS. Surprisingly, mammalian NO signalling and bacterial biosynthetic nitration share an evolutionary origin.

In mammals, the production of NO is catalysed solely by three highly regulated isoenzymes of NOS. NOSs produce NO from the oxidation of L-arginine (L-Arg) to L-citrulline through the intermediate N-hydroxy-L-Arg^{12–14}. Mammalian NOSs are homodimers that contain an amino-terminal haem oxygenase domain (NOS_{oxy}) and a carboxy-terminal flavoprotein reductase domain (NOS_{red}). The oxygenase domain binds L-Arg, haem and the redox-active cofactor 6R-tetrahydrobiopterin (H₄B), whereas the reductase domain binds FAD, FMN and NADPH. A calmodulin-binding sequence links the oxygenase and the reductase domains and regulates the reduction of NOS_{oxy} by NOS_{red} in those isoforms

that respond to Ca^{2+} .

Genome sequencing has revealed truncated NOS proteins in some Gram-positive bacteria, including *Deinococcus radiodurans*, *Staphylococcus aureus*, *Bacillus subtilis*, *B. halodurans* and *B. anthracis*. Bacterial NOSs are homologous to the mammalian NOS_{oxy} but lack an associated NOS_{red} and N-terminal regions that bind Zn^{2+} , the dihydroxypropyl side chain of H_4B , and the adjacent subunit of the dimer. Nevertheless, the *D. radiodurans* NOS and *B. subtilis* NOS are dimeric, have a haem liganded by a cysteine thiol group, bind substrate L-Arg and produce nitrogen oxide species in a manner dependent on pterin (either with H_4B or with the related cofactor tetrahydrofolate)^{9–11}. The crystal structure of *B. subtilis* NOS, complexed with L-Arg, confirmed that bacterial NOS proteins are similar to mammalian NOSs¹⁰, and NO production has been observed when a mammalian reductase domain is supplied¹¹. The redox mechanism by which the pterins H_4B and the analogue tetrahydrofolate support NO synthesis in bacterial NOSs mirrors that of the mammalian enzymes¹¹. A reductase protein that supplies electrons for bacterial NOS catalysis has not yet been identified, although comparisons of bacterial and mammalian NOS structures suggest a common mode of interaction for such a redox partner¹⁰.

Plant pathogenic *Streptomyces* species are the causal agents of potato scab disease, a globally important disease of potato. Pathogenicity depends on the production of a class of dipeptide phyto-toxins called thaxtomins⁸. While investigating the molecular genetics of plant pathogenicity in *Streptomyces* species, we discovered a gene with high sequence similarity to mammalian and bacterial NOSs. The location of this gene on a pathogenicity island that mobilizes among species to confer thaxtomin biosynthetic ability, and its proximity to two nonribosomal peptide synthases of the thaxtomin biosynthetic pathway, indicated that the NOS might participate in the nitration of thaxtomins (Fig. 1). The DNA sequences of the *nos* genes in *S. turgidiscabies*, *S. acidiscabies* and *S. scabies*, all of which produce thaxtomin A, are highly conserved (GenBank accession numbers AY204507–AY204509). Conservation of nearly all key residues in mammalian and *B. subtilis* NOSs that participate in cofactor binding, substrate binding and catalysis by *S. turgidiscabies* NOS (stNOS) suggests that stNOS is capable of producing NO from L-Arg (Supplementary Fig. S1). As in other bacterial NOSs, the reductase domain and the calmodulin-

binding site typical of mammalian NOSs are absent from stNOS. However, stNOS has an extended N-terminal region that other bacterial proteins lack (Supplementary Fig. S1). A mammalian NOS zinc-binding motif absent in the N termini of other bacterial NOSs might be conserved in some of the NOSs from pathogenic *Streptomyces* spp.

To determine whether stNOS is required for the production of thaxtomin A, *nos* was deleted from *S. turgidiscabies* (Fig. 2a, b). The Δnos strain produced only trace amounts of thaxtomin A ($0.20 \pm 0.02 \mu\text{g ml}^{-1}$; standard deviation reported, $n = 3$ for each strain tested) compared with the wild-type strain ($18.31 \pm 0.92 \mu\text{g ml}^{-1}$). Denitrothaxtomin was not detectable in the medium. Deletion of *nos* did not affect bacterial growth but did eliminate disease on potato tubers. Complementation of *nos* was achieved by expressing *nos* with the thiostrepton-inducible promoter on plasmid pIJ8600, which integrates into the chromosomal ϕC31 phage attachment site¹⁵ (Fig. 2b). The complemented strain Δnos pIJ8600-NOS increased the amount of thaxtomin A produced more than

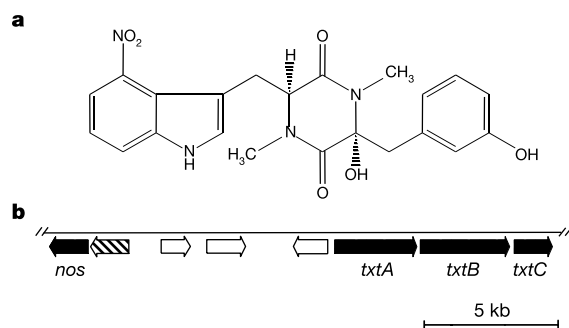


Figure 1 The *nos* gene is upstream of characterized thaxtomin biosynthetic genes. **a**, Thaxtomin A structure. Thaxtomin A is the predominant thaxtomin congener produced by plant-pathogenic *Streptomyces* spp. **b**, Genetic organization of the *nos* region in the thaxtomin-producing species *S. turgidiscabies*. The *txtAB* genes encode two similar peptide synthetases required for synthesis of the dipeptide⁸. The *txtC* gene encodes a P450 monooxygenase that is required for post-cyclization hydroxylation of the dipeptide³⁰. Black arrows, open reading frames involved in thaxtomin biosynthesis; hatched arrow, P450 monooxygenase homologue; open arrows, mobile genetic elements. kb, kilobases.

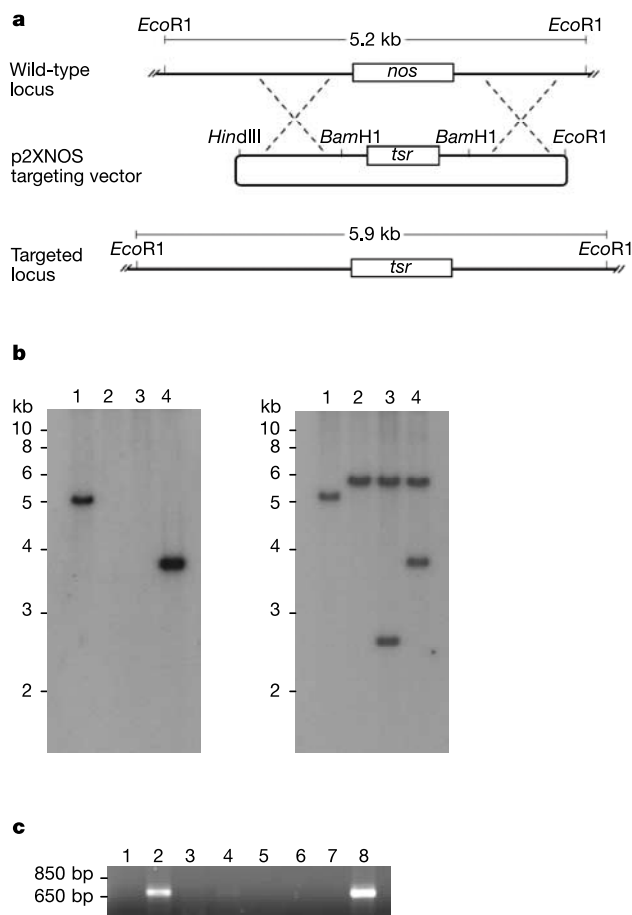


Figure 2 Deletion of *nos* decreases thaxtomin production by *S. turgidiscabies*. **a**, Model for *nos* replacement in *S. turgidiscabies* Car8 by the p2XNOS targeting vector with the use of marker-exchange mutagenesis. **b**, Southern hybridization of an internal 673 bp $\alpha\text{-}^{32}\text{P}$ -labelled *nos* probe (left) or the $\alpha\text{-}^{32}\text{P}$ -labelled 4.7 kb *HindIII/EcoRI* p2XNOS insert (right) to *EcoRI* digests of *S. turgidiscabies* wild-type (lane 1), Δnos (lane 2), Δnos pIJ8600 (empty vector) (lane 3), Δnos pIJ8600NOS (complemented vector) (lane 4) genomic DNA. The thiostrepton resistance gene (*tsr*) in the p2XNOS insert probe also hybridizes to *tsr* on pIJ8600 (lanes 3 and 4). **c**, Confirmation of *nos* expression using the polymerase chain reaction with reverse transcription. Odd-numbered lanes, DNase-treated RNA; even-numbered lanes, cDNA. Lanes 1 and 2, wild-type *S. turgidiscabies*; lanes 3 and 4, Δnos ; lanes 5 and 6, Δnos pIJ8600; lanes 7 and 8, Δnos pIJ8600NOS. bp, base pairs.

25-fold ($3.68 \pm 0.06 \mu\text{g ml}^{-1}$ thaxtomin A) compared with the empty vector control strain $\Delta\text{nos pIJ8600}$ ($0.13 \pm 0.01 \mu\text{g ml}^{-1}$). Expression of *nos* in the *S. turgidiscabies* wild-type strain was confirmed by using the polymerase chain reaction with reverse transcription (Fig. 2c). *nos* was not expressed in the Δnos or $\Delta\text{nos pIJ8600}$ strains, but expression was restored in the complemented strain $\Delta\text{nos pIJ8600NOS}$.

For stNOS to be responsible for thaxtomin nitration, stNOS should have activity characteristic of NO synthases. Furthermore, this activity and thaxtomin production should depend on L-arginine. We found that overexpressed stNOS in *Escherichia coli* produced nitrite from the NOS catalytic intermediate N-hydroxy-L-arginine by using a standard assay that measures nitrite production as an end product of NO or related species reacting in an oxygenated solution (Fig. 3a, b). NO synthase activity of the recombinant stNOS was inhibited by the selective NOS inhibitor N-nitro-L-arginine (Fig. 3b). We also evaluated NOS inhibitors for their effect on thaxtomin A production by *S. turgidiscabies*. Nitro-L-arginine methyl ester (NAME), a cell-permeable precursor of N-nitro-L-arginine, greatly suppressed thaxtomin production by *S. turgidiscabies* (IC_{50} 15 μM) but did not affect bacterial growth (Fig. 3c). Inhibition of thaxtomin production occurred at concentrations

commensurate with the binding affinities of NAME to mammalian NOSs^{12,16}. The other three NOS inhibitors tested also suppressed the production of thaxtomin A but to a smaller extent than NAME. Thus, inhibitors specific for the conserved NOS active centre curtail both NO synthase activity from recombinant stNOS *in vitro* and thaxtomin production *in vivo*.

Mammalian NOSs convert a terminal guanidino nitrogen of the L-Arg substrate to an oxidized nitrogen species. Our model for the role of stNOS in the nitration of thaxtomin predicts that the nitrate nitrogen derives from a terminal guanidino nitrogen of L-Arg. We conducted a ^{15}N feeding study: either L-Arg-guanidino- $^{15}\text{N}_2\cdot\text{HCl}$ or $^{15}\text{NH}_4\text{NO}_3$ was added to cultures of *S. turgidiscabies* under conditions that induce thaxtomin production. A ^{15}N NMR analysis of thaxtomin A extracted from cultures fed with the L-Arg-guanidino- $^{15}\text{N}_2\cdot\text{HCl}$ detected label only at the 4-nitro position, whereas in thaxtomin A extracted from control cultures that were fed with $^{15}\text{NH}_4\text{NO}_3$, all four nitrogens in thaxtomin were equally labelled (Fig. 4a, b). Electron-impact-ionization mass spectra of thaxtomin A also indicated specific incorporation of ^{15}N into the nitro group when the cultures were fed with L-Arg-guanidino- $^{15}\text{N}_2\cdot\text{HCl}$ (data not shown). We know of no enzymatic process other than NOS activity that converts the terminal guanidino nitrogen of L-Arg to an oxidized nitrogen species capable of nitration. These results provide definitive evidence for the role of stNOS in nitration of thaxtomin.

The specific nitration of a tryptophanyl moiety for biosynthesis is an unprecedented metabolic role for a NOS protein, although mammalian NOS production of nitroxyl (NO^-) and superoxide ($\text{O}_2^{\cdot-}$) is well characterized and such products could readily lead to nitrating species¹². Biosynthetic nitration reactions are rare and usually involve the oxidation of an amine¹⁷. The chemical mechanism of a NOS-mediated nitration might be complex because NO is unlikely to react directly with indole^{18–20}. Nevertheless, oxidized forms of NO such as nitrosonium (NO^+), peroxyxynitrite (ONOO^-), nitronium (NO_2^+) or nitrogen dioxide (NO_2) actively nitrate

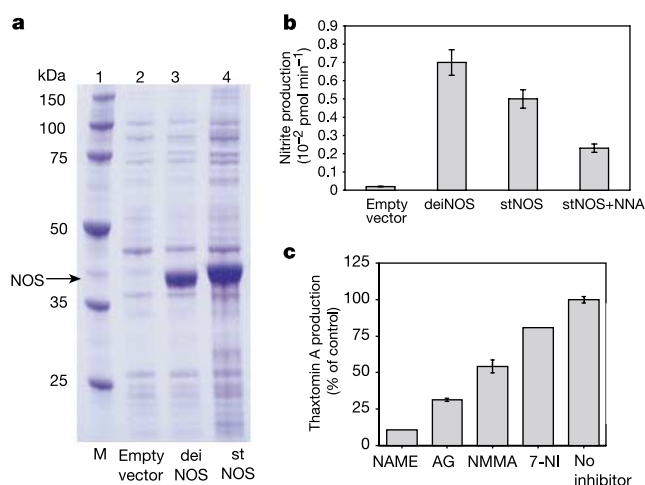


Figure 3 Nitrite formation from N^ω-hydroxy-L-arginine by recombinant stNOS, and sensitivity to inhibitors of mammalian NOSs. **a**, *S. turgidiscabies* NOS $\Delta 41$ (residues 41 to C terminus) was cloned into pet28 (Novagene) and expressed in *E. coli* BL-21 cells (lane 4). Full-length stNOS was rapidly degraded in *E. coli*, but the $\Delta 41$ construct, which contains all NOS homology regions, produced high concentrations of protein that were comparable to concentrations of recombinant *D. radiodurans* NOS¹⁰ (lane 3) and significantly above concentrations of background proteins expressed from cells containing the pet28 vector with no NOS insert (lane 2). Five microlitres of sonicated cell lysates from 3-ml cultures were resolved by 12% SDS–polyacrylamide-gel electrophoresis after 2.5 h induction with 0.1 mM isopropyl- β -D-thiogalactoside. The arrow shows the positions of recombinant *D. radiodurans* NOS (deiNOS) and stNOS, respectively, in lanes 3 and 4. M, molecular mass standards. **b**, Nitrite production by *E. coli* cell lysates depicted in **a**, measured by the Griess assay. The NOS intermediate N^ω-hydroxy-L-arginine (2 mM) and hydrogen peroxide (4 mM) were reacted with 20 μl of cell lysates shown in **a** and assayed for nitrite with the Griess reagents after 15 min. Lysates from cells overexpressing stNOS generate nitrite at concentrations comparable to that of overexpressed deiNOS and significantly above that of the empty vector control (pet28). The selective mammalian NOS inhibitor N-nitro-L-arginine (NNA) inhibits nitrite formation by lysates expressing stNOS when added to a concentration of 1 mM (same reaction conditions as above). **c**, The L-arginine-based NOS inhibitors nitro-L-arginine methyl ester (NAME, an NNA precursor, 32 μM), aminoguanidine (AG, 270 μM), N^G-monomethyl-L-arginine (NMMA, 510 μM) and 7-nitroindazole (7-NI, 1.2 μM) each inhibit toxin production by *S. turgidiscabies* without affecting bacterial growth (bars indicate standard error; $n = 6$).

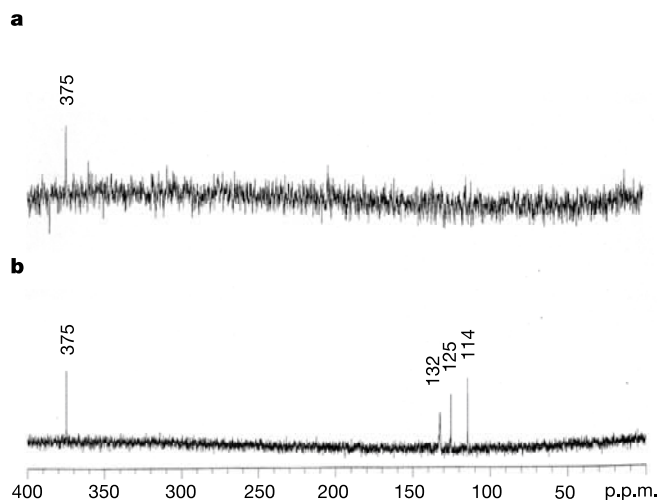


Figure 4 The nitrate nitrogen of thaxtomin A derives from the guanidine nitrogen of L-Arg. Shown are ^{15}N NMR spectra of thaxtomin A extracted from cultures fed with either **(a)** L-arginine-guanido- $^{15}\text{N}_2\cdot\text{HCl}$ or **(b)** $^{15}\text{NH}_4\text{NO}_3$. Long-range correlations between ^1H and ^{15}N chemical shifts in the $^{15}\text{NH}_4\text{NO}_3$ -labelled sample assign the two most upfield peaks at 125 and 114 p.p.m. to the N-methyl amide nitrogens, the peak at 132 p.p.m. to the indole nitrogen, and the downfield peak at 375 p.p.m. to the NO_2 -nitrogen. Accordingly, the single peak at 375 p.p.m. in the spectrum of the L-Arg-guanidino- $^{15}\text{N}_2\cdot\text{HCl}$ -supplied sample in **a** represents the signal of the NO_2 -nitrogen alone.

aromatic amino acids^{18–20}, and such reactions have been implicated in mammalian signalling^{21,22}. Given the high reactivities of these nitrogen oxide species towards tryptophan, the substrate nitrated probably contains indole; whether this substrate is tryptophan, an immediate precursor of thaxtomin, or some other intermediate is currently under investigation. Tryptophan nitration by peroxynitrite results primarily in modification at the indole 6-position²³. Thus, the production of a 4-nitrotryptophanyl moiety (Fig. 1a) suggests some enzymatic control of the nitration reaction. Nevertheless, minimal thaxtomin production in the Δnos strain indicates that diffusible nitration agents produced by other cellular processes can react productively with a thaxtomin precursor.

Conservation of residues in the haem pocket, the pterin site and the substrate access channel in the bacterial NOSs suggests a common function for prokaryotic enzymes and distinguishes them from their mammalian counterparts (Supplementary Fig. S1). A distinguishing catalytic feature between mammalian NOS and *B. subtilis* NOS is that the bacterial protein retains product NO coordinated to its haem iron 10–20-fold longer than do the mammalian enzymes¹¹. This correlates with a more sequestered immediate haem pocket that seems to be conserved in stNOS¹⁰. Perhaps a slower NO release allows bacterial proteins to direct NO reactivity by either sequestering NO close to the site of substrate binding or facilitating reaction with O₂. We suggest that other bacterial NOSs might also participate in the biosynthesis of secondary metabolites through the nitration of an amino acid or peptide substrate. Truncated NOSs have been discovered in only a subset of Gram-positive genomes and are absent from Gram-negative genomes, supporting a role for NOS proteins in secondary metabolism rather than a signalling function essential for vitality. Nitrated compounds are produced by other bacteria and fungi^{17,24,25} and at least one is likely to derive from an aromatic nitration reaction¹⁷. NOS-initiated nitrations would produce a class of compounds with unique chemical reactivity and diverse biological activities. From an evolutionary perspective, it is interesting that homologous enzymes are responsible for biosynthetic nitration reactions in bacteria and NO signalling in mammals. □

Methods

Thaxtomin production

S. turgidiscabies cultures were grown in oat bran broth or oat meal broth^{8,26} inoculated with spores. After shaking at 150 r.p.m. and 25 °C for 5–9 days the cultures were filtered and the dry weight of the mycelia was measured. Thaxtomin was extracted from the filtrate with ethyl acetate, dried, redissolved in methanol and quantified by HPLC (column: 5 μ m C₁₈, 250 mm $\times$ 4.6 mm; mobile phase: MeCN/water/trifluoroacetic acid (40:60:0.1, by volume)). In experiments investigating the suppression of thaxtomin production in the presence of L-NAME, this NOS inhibitor was dissolved in water, filter-sterilized and added at the time of inoculation.

¹⁵N feeding studies

NMR spectra were acquired for thaxtomin A extracted and purified from cultures fed with either L-Arg-guanidino-¹⁵N₂*HCl (Cambridge Isotope Laboratories) or ¹⁵NH₄NO₃ (Aldrich). ¹⁵NH₄NO₃ was added to oat bran broth at the time of inoculation with spores of *S. turgidiscabies*, whereas L-Arg-guanidino-¹⁵N₂*HCl was added just before the onset of thaxtomin biosynthesis (4–5 days after inoculation). ¹⁵N NMR spectra were acquired for thaxtomin A samples in C₂H₅O₂H at 40.5 MHz on a Varian VXR-400S spectrometer with a Nalorac broad-band probe affording observation of ¹⁵N at 40.5 MHz. Two-dimensional ¹H-¹⁵N long-range heteronuclear multiple bond correlation spectra with a long-range delay optimized for 8.3 Hz were acquired on a Varian Inova 500 spectrometer with an inverse detection triple resonance probe. Spectra were referenced externally by observing the ¹⁵N signal of formamide (90% solution in dimethyl sulphoxide) and then acquiring sample spectra with identical parameter sets. A conventional chemical shift scale (liquid NH₃ = 0 p.p.m.) was established by referencing the formamide signal to its reported value of 112 p.p.m.²⁷.

Molecular biology

DNA and RNA manipulation were performed with the use of standard techniques. Transformation of *S. turgidiscabies* was performed by using the polyethylene glycol (PEG)-mediated transformation of *S. turgidiscabies* protoplasts²⁸ with plasmid vectors propagated in *E. coli* ET12567²⁹. *S. turgidiscabies* p2XNOS single crossover transformants (apramycin resistant, thiostrepton resistant) were grown for three generations of

growth and sporulation on minimal-medium-containing thiostrepton, after which colonies containing double crossover (apramycin sensitive, thiostrepton resistant) recombination events were screened. Expression of *nos* was induced in the complemented Δnos strain by the addition of 10 μ g ml⁻¹ thiostrepton in a 5-ml oat bran culture 3 days after inoculation.

Received 3 January; accepted 19 March 2004; doi:10.1038/nature02504.

1. Lipton, S. Physiology. Nitric oxide and respiration. *Nature* **413**, 118–119 (2001).
2. Nathan, C. Inducible nitric oxide synthase: What difference does it make? *J. Clin. Invest.* **100**, 2417–2423 (1997).
3. Christopherson, K. S. & Bredt, D. S. Nitric oxide in excitable tissues: Physiological roles and disease. *J. Clin. Invest.* **100**, 2424–2429 (1997).
4. Garthwaite, J. & Boulton, C. L. Nitric-oxide signaling in the central-nervous system. *Annu. Rev. Physiol.* **57**, 683–706 (1995).
5. Torreilles, J. Nitric oxide: one of the more conserved and widespread signaling molecules. *Front. Biosci.* **6**, D1161–D1172 (2001).
6. Chen, Y. J. & Rosazza, J. P. N. Purification and characterization of nitric oxide synthase (NOSoc) from a *Nocardia* species. *J. Bacteriol.* **177**, 5122–5128 (1995).
7. Choi, W. S., Chang, M. S., Han, J. W., Hong, S. Y. & Lee, H. W. Identification of nitric oxide synthase in *Staphylococcus aureus*. *Biochem. Biophys. Res. Commun.* **237**, 554–558 (1997).
8. Healy, F. G., Wach, M., Krasnoff, S. B., Gibson, D. M. & Loria, R. The *txtAB* genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol. Microbiol.* **38**, 794–804 (2000).
9. Adak, S. *et al.* Cloning, expression, and characterization of a nitric oxide synthase protein from *Deinococcus radiodurans*. *Proc. Natl Acad. Sci. USA* **99**, 107–112 (2002).
10. Pant, K., Bilwes, A. M., Adak, S., Stuehr, D. J. & Crane, B. R. Structure of a nitric oxide synthase heme protein from *Bacillus subtilis*. *Biochemistry* **41**, 11071–11079 (2002).
11. Adak, S., Aulak, K. & Stuehr, D. J. Evidence for nitric oxide production by a nitric-oxide synthase-like protein from *Bacillus subtilis*. *J. Biol. Chem.* **277**, 16167–16171 (2002).
12. Alderton, W. K., Cooper, C. E. & Knowles, R. G. Nitric oxide synthases: structure, function and inhibition. *Biochem. J.* **357**, 593–615 (2001).
13. Pfeiffer, S., Mayer, B. & Hemmens, B. Nitric oxide: Chemical puzzles posed by a biological messenger. *Angew. Chem. Int. Edn Engl.* **38**, 1714–1731 (1999).
14. Stuehr, D. J. Mammalian nitric oxide synthases. *Biochim. Biophys. Acta* **1411**, 217–230 (1999).
15. Sun, J., Kelemen, G. H., Fernandez-Abalos, J. M. & Bibb, M. J. Green fluorescent protein as a reporter for spatial and temporal gene expression in *Streptomyces coelicolor* A3(2). *Microbiology* **145**, 2221–2227 (1999).
16. Southan, G. J. & Szabo, C. Selective pharmacological inhibition of distinct nitric oxide synthase isoforms. *Biochem. Pharmacol.* **51**, 383–394 (1996).
17. Carter, G. T. *et al.* Direct biochemical nitration in the biosynthesis of dioxapyrrolomycin: A unique mechanism for the introduction of nitro groups in microbial products. *J. Chem. Soc. Chem. Commun.* **17**, 1271–1273 (1989).
18. Koppenol, W. H. The basic chemistry of nitrogen monoxide and peroxynitrite. *Free Radical Biol. Med.* **25**, 385–391 (1998).
19. Hughes, M. N. Relationships between nitric oxide, nitroxyl ion, nitrosonium cation and peroxynitrite. *Biochim. Biophys. Acta* **1411**, 263–272 (1999).
20. Ridd, J. H. Some unconventional pathways in aromatic nitration. *Acta Chem. Scand. A* **52**, 11–22 (1998).
21. Packer, L. E. (ed.) Nitric oxide. Part B. Physiological and pathological processes. *Methods Enzymol.* **268** (1996).
22. Ischiropoulos, H. (ed.) Serial review: Reactive nitrogen species, tyrosine nitration and cell signaling. *Free Radical Biol. Med.* **33** (2002).
23. Alvarez, B. *et al.* Peroxynitrite-dependent tryptophan nitration. *Chem. Res. Toxicol.* **9**, 390–396 (1996).
24. Jalal, M. A. F., Hossain, M. B. & Van der Helm, D. Structure of anticancer antibiotic L-alanosine. *Acta Crystallogr. C* **42**, 733–738 (1986).
25. Ohba, K. *et al.* Nitropeptin, a new dipeptide antibiotic possessing a nitro group. *J. Antibiot.* **40**, 709–713 (1987).
26. Goyer, C., Vachon, J. & Beaulieu, C. Pathogenicity of *Streptomyces scabies* mutants altered in thaxtomin A production. *Phytopathology* **88**, 442–445 (1998).
27. Martin, G. E. & Hadden, C. E. Long-range ¹H-¹⁵N heteronuclear shift correlation at natural abundance. *J. Nat. Prod.* **63**, 543–585 (2000).
28. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Streptomyces Genetics* (John Innes Foundation, Norwich, 2000).
29. MacNeil, D. J. *et al.* Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61–68 (1992).
30. Healy, F. G., Krasnoff, S. B., Wach, M., Gibson, D. M. & Loria, R. Involvement of a cytochrome P450 monooxygenase in thaxtomin A biosynthesis by *Streptomyces acidiscabies*. *J. Bacteriol.* **184**, 2019–2029 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Condo Jr and I. Keresztes of the NMR facility of the Department of Chemistry and Chemical Biology, Cornell University, and staff of the Mass Spectrometry Laboratory, School of Chemical Sciences, University of Illinois Urbana-Champaign, for assistance in acquiring spectra for isotopic labelling studies. We also thank R. Parry for helpful discussion, and K. Pant and Madhavan Buddha for help with the NOS assays. We acknowledge the support of the National Science Foundation and the United States Department of Agriculture National Research Initiative, Biology of Plant–Microbe Associations Program.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.L. (rl21@cornell.edu).

Mechanotransduction through growth-factor shedding into the extracellular space

Daniel J. Tschumperlin¹, Guohao Dai³, Ivan V. Maly³, Tadashi Kikuchi², Lily H. Laiho³, Anna K. McVittie³, Kathleen J. Haley², Craig M. Lilly², Peter T. C. So³, Douglas A. Lauffenburger³, Roger D. Kamm³ & Jeffrey M. Drazen¹

¹Physiology Program, Department of Environmental Health, Harvard School of Public Health, and ²Pulmonary and Critical Care Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

³Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Physical forces elicit biochemical signalling in a diverse array of cells, tissues and organisms^{1–3}, helping to govern fundamental biological processes. Several hypotheses have been advanced that link physical forces to intracellular signalling pathways, but in many cases the molecular mechanisms of mechanotransduction remain elusive^{1–9}. Here we find that compressive stress shrinks the lateral intercellular space surrounding epithelial cells, and triggers cellular signalling via autocrine binding of epidermal growth factor family ligands to the epidermal growth factor receptor. Mathematical analysis predicts that constant rate shedding of autocrine ligands into a collapsing lateral intercellular space leads to increased local ligand concentrations that are sufficient to account for the observed receptor signalling; direct experimental comparison of signalling stimulated by compressive stress versus exogenous soluble ligand supports this prediction. These findings establish a mechanism by which mechanotransduction arises from an autocrine ligand–receptor circuit operating in a dynamically regulated extracellular volume, not requiring induction of force-dependent biochemical processes within the cell or cell membrane.

Epithelial and endothelial cells line the surfaces of the body and

frequently encounter physical stresses as pressures equilibrate across these cellular barriers^{10–13}. We studied the morphological and biochemical responses of normal human bronchial epithelial (NHBE) cells to compressive stress: *in vivo* they experience such stress as a consequence of airway constriction¹⁴, and *in vitro* compressive stress triggers an integrated pro-fibrotic response reminiscent of that seen in chronic asthma^{15,16}.

We used two-photon microscopy to visualize the morphology of living NHBE cells both before and during the application of physical stress. Cell-permeant cytoplasmic dyes (Celltracker and calcein AM) were used to visualize the cell bodies, and a cell-impermeant fluorophore-conjugated dextran was used to visualize the extracellular space directly. The cells, which were grown at the air–liquid interface to produce a pseudo-stratified, mucociliary epithelium, were joined by tight junctions at their apical surface, and separated along their basolateral surface by a lateral intercellular space (LIS) comprising $15 \pm 3\%$ (mean $\pm$ s.d., $n = 5$) of the total tissue volume. This morphology, including the LIS, is typical of stratified epithelium¹⁷ and is representative of the airways *in vivo*^{18,19}.

To expose the cells to a compressive stress similar to that resulting from airway constriction we applied an apical to basal transcellular pressure gradient across the epithelium^{14,16}. This compressive stress decreased the thickness of NHBE cell layers an average of $10.6 \pm 1.9\%$ (mean $\pm$ s.d., $n = 7$, Supplementary Table 1). However, total cell volume was not significantly altered by the compressive stress (decrease of $1.6 \pm 2.3\%$, $n = 7$, Supplementary Table 1), suggesting that changes in cell layer height were a result of fluid leaving the LIS. This was confirmed qualitatively by visualizing the extracellular space (Fig. 1). Quantitative determination of LIS volume in tissue specimens before and after exposure to compressive stress demonstrated an $87.3 \pm 10.0\%$ reduction in LIS volume (mean $\pm$ s.d., $n = 5$, Supplementary Table 2). Our findings indicate that although epithelial cell volumes are maintained relatively constant under compressive stress, the LIS is highly compliant and susceptible to large percentage volume changes.

Collapse of the LIS under compressive stress brings the lateral surfaces of adjacent cells closer to one another. The basolateral surfaces of NHBE cells exhibit strong staining for the ErbB1 epidermal growth factor receptor (EGFR, Fig. 2a), in agreement

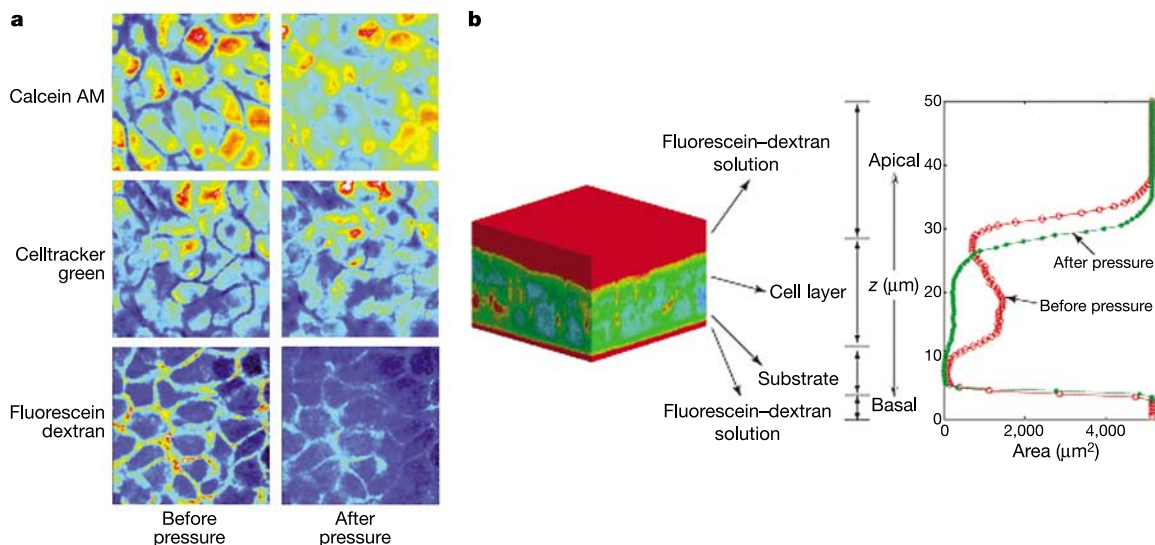


Figure 1 Compliance of the lateral intercellular space. **a**, Pseudo-colour x–y slices from identical mid-cell layers obtained by two-photon microscopy before and during application of 20 cm H₂O transcellular pressure. **b**, A typical pseudo-colour three-dimensional reconstruction of bronchial epithelial cells negatively stained with cell-impermeant fluorescein–dextran. Red corresponds to dextran, whereas cooler colours

(blue–green) correspond to the absence of dextran. The intensity of dextran staining is plotted along the z axis. The red contour indicates the profile of dextran staining under control conditions, whereas the green profile was obtained in the same tissue 30 min after the onset of a continuous 20 cm H₂O transcellular pressure. Note the decrease in cell layer height and loss of LIS volume.

with observations made in intact human airways and primary airway epithelial cells^{20,21}. As previously shown²², application of compressive stress for 30 min elicited a robust increase in phosphorylation of the mitogen-activated protein kinase ERK (Fig. 2b). This response was preceded by an increase in phosphorylation of the EGFR at 5–20 min (Fig. 2a). Blockade of the tyrosine kinase domain of the EGFR with basolaterally applied tyrphostin AG 1478 strongly attenuated the mechanical stress-induced phosphorylation of ERK, whereas incubation with tyrphostin AG 1296, which inhibits the tyrosine kinase domain of the platelet-derived growth factor receptor, had no effect on the mechanically induced signal (Fig. 2b). Basolateral application of a monoclonal function-blocking EGFR antibody before compressive stress attenuated the compressive stress-induced ERK phosphorylation in a concentration-dependent manner (Fig. 2b), confirming a role for the extracellular ligand-binding domain of the EGFR. To identify the EGF family member responsible, we applied compressive stress in the presence of increasing basolateral amounts of neutralizing antibodies to EGF, transforming growth factor- α (TGF- α) and heparin-binding epidermal growth factor (HB-EGF). Whereas the EGF and TGF- α antibodies were ineffective at the highest concentration used, the antibody to HB-EGF attenuated the compressive stress-induced ERK phosphorylation in a dose-dependent manner (Fig. 2b). Incubation with GM 6001, a broad-spectrum inhibitor of matrix metalloprotease activity, was used to inhibit ectodomain shedding of transmembrane EGF family precursors from the basolateral cell surface^{23,24}. Its presence attenuated both baseline and transcellular pressure-induced increases in ERK phosphorylation, demonstrat-

ing that constitutive sheddase activity for EGF family ligands is present in the LIS, and is required for transduction of the applied mechanical stress. These results demonstrate that compressive stress elicits ERK phosphorylation through a LIS-restricted autocrine pathway involving metalloprotease-dependent shedding of HB-EGF, leading to subsequent binding and activation of the EGFR.

To confirm the relevance of this signalling response in the native setting of the airway wall, we fixed isolated mouse lungs after administration of various bronchially active substances, and examined the distribution of epithelial staining for the active (phosphorylated) form of the EGFR. A 5 min tracheal perfusion with methacholine (MCh; 1 mM), which was sufficient to maximally constrict the airways, significantly increased the number of airways positive for the phosphorylated EGFR (Fig. 2c). This response to MCh was specific to the mechanical stress evoked by muscle constriction, as pre-treatment with isoproterenol (Isp), which greatly attenuated constriction, abrogated the EGFR response to MCh.

The localization of the EGFR autocrine loop at the perimeter of the collapsing LIS suggests a proximal role for this receptor system in the transduction of mechanical stress in epithelial cells. We considered the possibility that transient shear stresses generated as fluid is squeezed out of the collapsing LIS might trigger EGFR activation; however, when we applied transcellular pressure gradually as a ramp function over 30 min to minimize intercellular shear stress, signal transduction was stimulated as effectively as by a step increase in transcellular pressure (Supplementary Fig. 1). We also considered the possibility that the compressive stress response

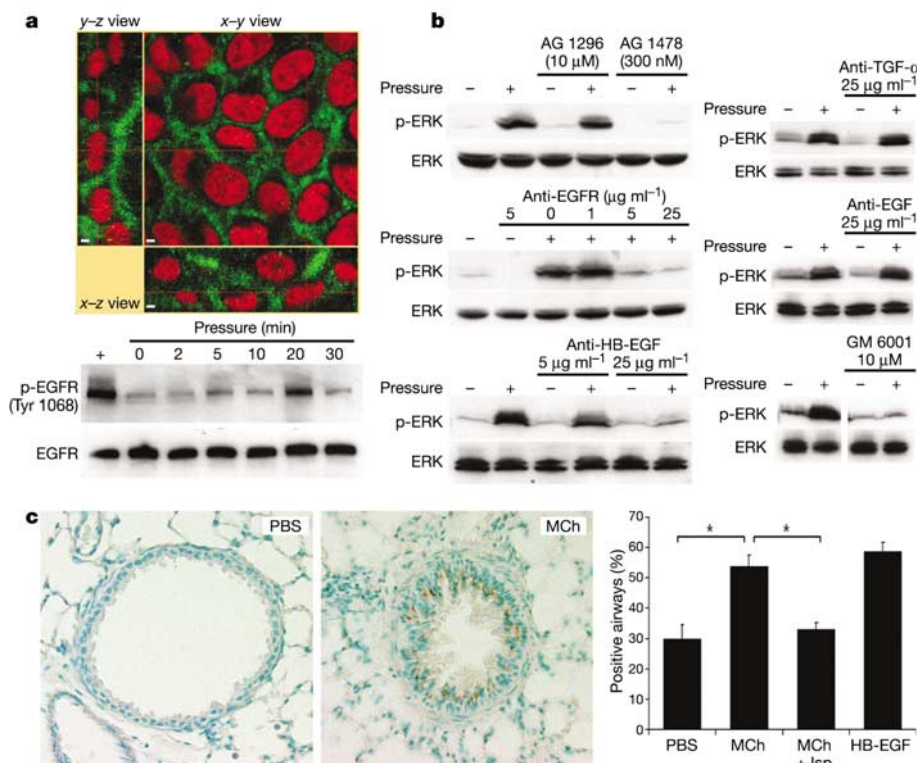


Figure 2 The EGFR mediates transduction of compressive stress. **a**, NHBE cells exhibit basolaterally polarized distribution of EGFR (green) lining the LIS; cell nuclei are counterstained with propidium iodide (red). Scale bars indicate 2 μ m. The EGFR is rapidly phosphorylated (Tyr 1068) in response to transcellular compressive stress (pressure). **b**, NHBE cells exhibit ERK phosphorylation after 30 min exposure to transcellular pressure (20–30 cm H₂O); this response is attenuated in the presence of the EGFR inhibitor AG 1478 but not the platelet-derived growth factor receptor inhibitor AG 1296, and is also attenuated by a function-blocking antibody that prevents ligand binding to the EGFR, a neutralizing antibody against HB-EGF, and the matrix metalloprotease inhibitor GM 6001,

but not neutralizing antibodies against EGF and TGF- α . All blots are representative of at least two independent experiments. **c**, Representative images show the increased immunostaining (brown) for the phosphorylated EGFR (Tyr 1068) in mouse lungs perfused for 5 min with MCh (1 mM) relative to those perfused with PBS. The increased staining (percentage positive airways; asterisk = $P < 0.0005$, one-factorial analysis of variance and Bonferroni/Dunn test, $n = 4$ –5, mean $\pm$ s.e.m.) was reversed by pre-treatment with isoproterenol (MCh + Isp), which abrogated the airway constriction. Perfusion with 10 ng ml⁻¹ HB-EGF is included as a positive control.

might result from EGFR transactivation secondary to G-protein-coupled receptor activation^{6,11,24} or ATP release^{8,25,26}. Pre-treatment with either pertussis toxin or the extracellular ATP scavenger apyrase was incapable of attenuating the ERK phosphorylation response to compression (Supplementary Fig. 1).

We next considered whether it is feasible that the geometric alterations in the LIS are sufficient to stimulate EGFR signalling by modulating ligand availability and receptor occupancy^{27–29}. EGF family ligands that are shed into LIS undergo molecular diffusion near the cell surface until either they are bound by cell-surface receptors, or they escape into the basal medium²⁷. To describe the steady-state concentration profile of ligand molecules in the LIS, and the dependence of this concentration profile on LIS geometry, we solved the mass balance equation for ligand shedding and diffusion, leading to the expression $c = c_m + (h^2 - X^2)r/(Dw)$. The concentration c in the LIS exceeds the concentration c_m in the medium, or in the underlying tissue *in vivo*, owing to the autocrine ligand release rate r . The epithelium allows outflux of the released ligand to the medium through the basal LIS opening but not across the tight junctions at the apical surface of the cells, and so the concentration depends on the distance X measured from the apex of the LIS (Fig. 3a). Assuming a uniform distribution of ligand shedding, constant in time, the concentration of ligands is highest adjacent to the tight junctions and decreases quadratically, reaching c_m at the basal LIS opening, where $X = h$, the LIS height. The difference in ligand concentration from that in the basal medium is proportional to the shedding rate r and inversely proportional to the diffusion constant of the ligand, D . Notably, we find that the ligand concentration is inversely proportional to the LIS width w . Thus, our analysis predicts that the effective ligand concentration to which cells are exposed is strongly dependent on the geometry of the LIS, which is itself a function of the mechanical environment of the

tissue. On the basis of this analysis we directly compared compressive stress to stimulation with soluble HB-EGF. In our imaging studies we measured an $87 \pm 10\%$ decrease in LIS volume, which should yield an approximately eightfold increase in HB-EGF concentration. Compressive stress produced a change in ERK phosphorylation equivalent to a roughly tenfold increase in HB-EGF concentration (Fig. 3b), consistent with our prediction.

Our findings support the hypothesis that cells can sense mechanical stress through autocrine loops localized to compliant extracellular spaces; this mechanism seems to be capable of acting both independently of other known mechanosensitive responses and in concert with other mechanisms to broaden and amplify the response. Spatially limited autocrine loops occur in a wide variety of cell types, especially in epithelia where they are frequently localized in the LIS. Our results emphasize the high relative compliance of the LIS, and indicate that cells respond to local changes in ligand availability as a result of geometric alterations in their intercellular spaces. A unique aspect of this mechanochemical transduction mechanism is that it requires no direct conformational or metabolic alteration in any protein by mechanical stress. We speculate that such a transduction system could be tuned to perceive a wide range of deformations and stresses, and may be important in the sensation of and responses to physical forces in a number of biological systems. □

Methods

Cell culture

NHBE cells (Clonetics-BioWhittaker) were cultured as previously described on uncoated microporous polyester inserts (Transwell-Clear, 0.4- μ m pore size, Costar)²². Experiments were carried out in minimal medium with no serum or growth factors except insulin ($5.7 \mu\text{g ml}^{-1}$) and transferrin ($5 \mu\text{g ml}^{-1}$). Function-blocking antibodies (all from R&D Systems except anti-EGFR clone LA1; Upstate Biotechnology), GM 6001, AG 1478, AG 1296 (all from Calbiochem), and soluble HB-EGF (R&D Systems) were applied to the basal surface of cultures only.

Two-photon imaging

NHBE cells were stained apically and basally with the fluorescent probes calcein AM ($10 \mu\text{M}$), Celltracker green ($10 \mu\text{M}$), or 70 kDa fluorescein-dextran ($100 \mu\text{M}$, all from Molecular Probes) diluted in culture medium. After incubation for 30 min each tissue culture insert was mounted on a custom-built two-photon microscope, and a silicon plug was used to create a pressure chamber over the apical surface of the cells. After a complete sequence of images was obtained under control conditions, an apical-to-basal transcellular pressure difference of 20 cm H_2O was applied. Imaging was then repeated within the same region of cells. Images were acquired at intervals of 0.5 μm in the z direction, with each image acquired at 256×256 pixels, 0.28 μm pixel dimension. Image stacks were analysed using Spyglass Slicer (Spyglass).

Western blotting

Cell lysates were separated in 10% precast polyacrylamide gels (BioRad)²². Proteins were transferred to polyvinylidene difluoride membranes (Immobilon-P, Millipore), blocked with 5% milk, and incubated overnight (4°C) in primary antibodies against ERK, phospho-ERK, EGFR, or phospho-EGFR (Tyr 1068; all from Cell Signaling Technology). Bands were visualized with horseradish peroxidase-conjugated anti-rabbit IgG (1:2,000) and enhanced chemiluminescence.

Immunofluorescence

Cells were fixed with 3% paraformaldehyde and stained with a mouse monoclonal antibody to the extracellular domain of the human EGFR (clone LA1, Upstate Biotechnology) and a secondary goat anti-mouse IgG antibody conjugated to Alexa Fluor 488 (Molecular Probes). Cell nuclei were counterstained with propidium iodide. Three-dimensional stacks of confocal images were obtained using a Sarastro 2000 confocal laser-scanning microscope.

Isolated lung perfusion

Lungs of male BALB/C mice (age 8 weeks) were tracheally perfused following the method of ref. 30 with PBS at a flow rate of 20 ml h^{-1} . Perfusate exited the lung via small holes (5 per lobe) poked in the pleura (27 gauge needle). The airway opening pressure (P_{ao}) at the trachea was continuously measured as an index of airway constriction. Perfusion with MCh induced a transient, concentration-dependent increase in P_{ao} , with a maximal response at 1 mM ($\Delta P_{ao} = 13 \text{ cm H}_2\text{O}$ at 5 min). Perfusion with 10^{-4} M Isp 1 min before and continuous with MCh perfusion attenuated MCh-induced bronchial constriction by 78%. Perfusion with Isp alone had little effect on P_{ao} . Perfused lungs from each group ($n = 4-5$) were fixed by perfusion at 5 min with 4% paraformaldehyde.

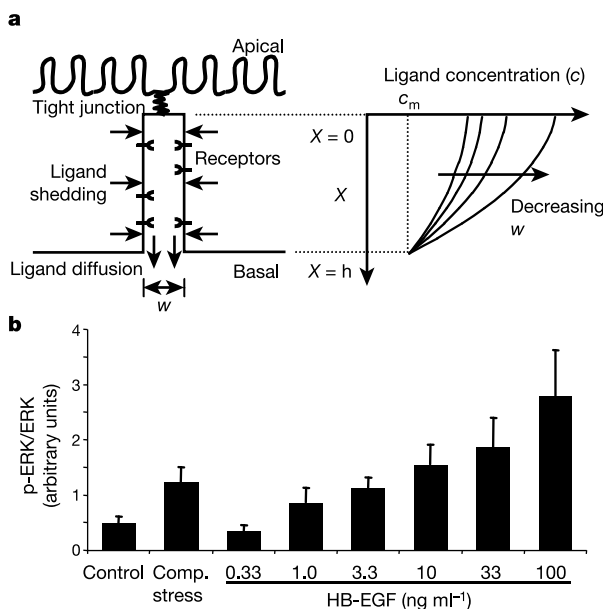


Figure 3 LIS width regulates autocrine signalling. **a**, Schematic of the epithelial LIS showing the boundary conditions regulating EGF family ligand diffusion, and example profiles of ligand concentration within the LIS. **b**, ERK phosphorylation elicited by compressive stress compared to 3.3-fold increasing concentrations of HB-EGF (mean $\pm$ s.e.m., $n = 5$). Concentrations of HB-EGF $< 0.33 \text{ ng ml}^{-1}$ had no effect on ERK phosphorylation. We thus assume that the baseline local concentration of EGF family ligands in the LIS is between 0.33 and 1.0 ng ml^{-1} . From equation (7) we predict that the result of compressive stress should be most comparable to the effect of adding exogenous HB-EGF to increase this concentration approximately tenfold, to between 3.3 and 10 ng ml^{-1} .

Immunohistochemistry

Paraffin-embedded lung sections were incubated with a rabbit polyclonal antibody against phospho-EGF receptor (Tyr 1068, Cell Signalling Technology); immunopositive areas were visualized with an avidin-biotin method using biotinylated goat anti-rabbit IgG secondary antibody (Vector Laboratories), the chromagen diaminobenzidine and methyl green counterstaining. Irrelevant IgG antibodies substituted for the primary antiserum were the negative controls. All cross-sectioned airways in each section were counted, and airways with any staining in the epithelium were counted as phospho-EGFR positive.

Mass balance analysis of the LIS

We approximate the geometry of the LIS as a two-dimensional gap of width w and height h between two surfaces (Fig. 3a), and assume that the rate of ligand shedding r , width w , and diffusion coefficient D are independent of position X along the apical-basal axis. Transport by convection is ignored. At steady state the rate of ligand entering the LIS via shedding from the cells bounding the LIS will be exactly balanced by the ligand leaving the LIS by diffusion through the basal opening of the LIS into the bulk medium. If we examine the mass transfer in an infinitesimal slice of LIS of width w and thickness dX at position X , the flux into the slice can be written as:

$$\text{influx} = 2rdX \quad (1)$$

where r is the density of ligand shedding per unit length, dX is the slice thickness, and the factor 2 comes from the two surfaces bounding the LIS in our idealized geometry. This influx is balanced by the net outflux due to diffusion, which is proportional to the spatial derivative of the ligand concentration gradient:

$$\text{outflux} = -\frac{d}{dX} \left[wD \frac{dc}{dX} \right] \quad (2)$$

where c is the concentration averaged over y . Equating influx and outflux, assuming constant r , w and D , and integrating from 0 to X , yields:

$$2r \int_0^X dX = -wD \int_0^X \frac{d}{dX} \left[\frac{dc}{dX} \right] \quad (3)$$

Our boundary conditions are zero flux through the impermeable tight junction bounding the apical end of the LIS ($X = 0$), and ligand concentration equilibrates with the bulk medium (c_m) at the basal opening of the LIS ($X = h$):

$$\left. \frac{dc}{dX} \right|_{X=0} = 0 \quad (4)$$

and

$$c|_{X=h} = c_m \quad (5)$$

Integrating equation (3) and using the boundary condition at $X = 0$ (equation (4)), the following relation is obtained:

$$\frac{dc}{dX} = -\frac{2rX}{wD} \quad (6)$$

Solving this equation for $c(X)$ and satisfying the second boundary condition (equation (5)) yields:

$$c = c_m + \frac{r}{wD} (h^2 - X^2) \quad (7)$$

Received 2 February; accepted 5 April 2004; doi:10.1038/nature02543.

Published online 21 April 2004.

- Blount, P. Molecular mechanisms of mechanosensation: big lessons from small cells. *Neuron* **37**, 731–734 (2003).
- Torday, J. S. & Rehan, V. K. Mechanotransduction determines the structure and function of lung and bone: a theoretical model for the pathophysiology of chronic disease. *Cell Biochem. Biophys.* **37**, 235–246 (2003).
- Farge, E. Mechanical induction of Twist in the *Drosophila* foregut/stomodaeal primordium. *Curr. Biol.* **13**, 1365–1377 (2003).
- Alenghat, F. J. & Ingber, D. E. Mechanotransduction: all signals point to cytoskeleton, matrix, and integrins. *Sci. STKE* **2002**, PE6 (2002).
- Hamill, O. P. & Martinac, B. Molecular basis of mechanotransduction in living cells. *Physiol. Rev.* **81**, 685–740 (2001).
- Gudi, S. R., Clark, C. B. & Frangos, J. A. Fluid flow rapidly activates G proteins in human endothelial cells. Involvement of G proteins in mechanochemical signal transduction. *Circ. Res.* **79**, 834–839 (1996).
- Shyy, J. Y. & Chien, S. Role of integrins in endothelial mechanosensing of shear stress. *Circ. Res.* **91**, 769–775 (2002).
- Dull, R. O., Tarbell, J. M. & Davies, P. F. Mechanisms of flow-mediated signal transduction in endothelial cells: kinetics of ATP surface concentrations. *J. Vasc. Res.* **29**, 410–419 (1992).
- Ali, M. H. & Schumacker, P. T. Endothelial responses to mechanical stress: where is the mechanosensor? *Crit. Care Med.* **30**, S198–S206 (2002).
- Han, O., Li, G. D., Sumpio, B. E. & Basson, M. D. Strain induces Caco-2 intestinal epithelial proliferation and differentiation via PKC and tyrosine kinase signals. *Am. J. Physiol.* **275**, G534–G541 (1998).
- Correa-Meyer, E., Pesce, L., Guerrero, C. & Sznajder, J. I. Cyclic stretch activates ERK1/2 via G proteins and EGFR in alveolar epithelial cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **282**, L883–L891 (2002).
- Ressler, B., Lee, R. T., Randell, S. H., Drazen, J. M. & Kamm, R. D. Molecular responses of rat tracheal epithelial cells to transmembrane pressure. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **278**, L1264–L1272 (2000).
- Lehoux, S. & Tedgui, A. Signal transduction of mechanical stresses in the vascular wall. *Hypertension* **32**, 338–345 (1998).
- Wiggs, B. R., Hrousis, C. A., Drazen, J. M. & Kamm, R. D. On the mechanism of mucosal folding in normal and asthmatic airways. *J. Appl. Physiol.* **83**, 1814–1821 (1997).

- Tschumperlin, D. J., Shively, J. D., Kikuchi, T. & Drazen, J. M. Mechanical stress triggers selective release of fibrotic mediators from bronchial epithelium. *Am. J. Respir. Cell Mol. Biol.* **28**, 142–149 (2003).
- Swartz, M. A., Tschumperlin, D. J., Kamm, R. D. & Drazen, J. M. Mechanical stress is communicated between different cell types to elicit matrix remodeling. *Proc. Natl Acad. Sci. USA* **98**, 6180–6185 (2001).
- Spring, K. R. & Hope, A. Size and shape of the lateral intercellular spaces in a living epithelium. *Science* **200**, 54–58 (1978).
- Ohashi, Y., Motojima, S., Fukuda, T. & Makino, S. Airway hyperresponsiveness, increased intracellular spaces of bronchial epithelium, and increased infiltration of eosinophils and lymphocytes in bronchial mucosa in asthma. *Am. Rev. Respir. Dis.* **145**, 1469–1476 (1992).
- Kondo, M., Finkbeiner, W. E. & Widdicombe, J. H. Changes in permeability of dog tracheal epithelium in response to hydrostatic pressure. *Am. J. Physiol.* **262**, L176–L182 (1992).
- Polosa, R. *et al.* Expression of c-erbB receptors and ligands in human bronchial mucosa. *Am. J. Respir. Cell Mol. Biol.* **20**, 914–923 (1999).
- Vermeer, P. D. *et al.* Segregation of receptor and ligand regulates activation of epithelial growth factor receptor. *Nature* **422**, 322–326 (2003).
- Tschumperlin, D. J. *et al.* Bronchial epithelial compression regulates MAP kinase signaling and HB-EGF-like growth factor expression. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **282**, L904–L911 (2002).
- Dong, J. *et al.* Metalloprotease-mediated ligand release regulates autocrine signaling through the epidermal growth factor receptor. *Proc. Natl Acad. Sci. USA* **96**, 6235–6240 (1999).
- Prenzel, N. *et al.* EGF receptor transactivation by G-protein-coupled receptors requires metalloproteinase cleavage of proHB-EGF. *Nature* **402**, 884–888 (1999).
- Homolya, L., Steinberg, T. H. & Boucher, R. C. Cell to cell communication in response to mechanical stress via bilateral release of ATP and UTP in polarized epithelia. *J. Cell Biol.* **150**, 1349–1360 (2000).
- Yamamoto, K. *et al.* Endogenously released ATP mediates shear stress-induced Ca^{2+} influx into pulmonary artery endothelial cells. *Am. J. Physiol. Heart Circ. Physiol.* **285**, H793–H803 (2003).
- Shvartsman, S. Y., Wiley, H. S., Deen, W. M. & Lauffenburger, D. A. Spatial range of autocrine signaling: modeling and computational analysis. *Biophys. J.* **81**, 1854–1867 (2001).
- Lauffenburger, D. A., Oehrman, G. T., Walker, L. & Wiley, H. S. Real-time quantitative measurement of autocrine ligand binding indicates that autocrine loops are spatially localized. *Proc. Natl Acad. Sci. USA* **95**, 15368–15373 (1998).
- Shvartsman, S. Y. *et al.* Autocrine loops with positive feedback enable context-dependent cell signaling. *Am. J. Physiol. Cell Physiol.* **282**, C545–C559 (2002).
- Martins, M. A., Shore, S. A., Gerard, N. P., Gerard, C. & Drazen, J. M. Peptidase modulation of the pulmonary effects of tachykinins in tracheal superfused guinea pig lungs. *J. Clin. Invest.* **85**, 170–176 (1990).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Shively and J. Foley for technical assistance, G. Horsley for early work on EGFR signalling, and J. Lai for performing conventional confocal imaging. This work was supported by funding from the NHLBI (J.M.D. and R.D.K.), NIGMS (D.A.L.) and a Parker B. Francis Fellowship in Pulmonary Research (D.J.T.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M.D. (jdrazen@nejm.org).

The ubiquitin ligase COP1 is a critical negative regulator of p53

David Dornan¹, Ingrid Wertz^{1,5}, Harumi Shimizu¹, David Arnott², Gretchen D. Frantz³, Patrick Dowd⁴, Karen O' Rourke¹, Hartmut Koeppen³ & Vishva M. Dixit¹

¹Department of Molecular Oncology, ²Department of Protein Chemistry,

³Department of Pathology and ⁴Department of Molecular Biology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA

⁵Department of Human Physiology, University of California, Davis, California 95616, USA

COP1 (constitutively photomorphogenic 1) is a RING-finger-containing protein that functions to repress plant photomorphogenesis, the light-mediated programme of plant development. Mutants of COP1 are constitutively photomorphogenic, and this has been attributed to their inability to negatively regulate the proteins LAF1 (ref. 1) and HY5 (ref. 2). The role of COP1 in mammalian cells is less well characterized³. Here we identify the tumour-suppressor protein p53 as a COP1-interacting protein. COP1 increases p53 turnover by targeting it for

degradation by the proteasome in a ubiquitin-dependent fashion, independently of MDM2 or Pirh2, which are known to interact with and negatively regulate p53. Moreover, COP1 serves as an E3 ubiquitin ligase for p53 *in vitro* and *in vivo*, and inhibits p53-dependent transcription and apoptosis. Depletion of COP1 by short interfering RNA (siRNA) stabilizes p53 and arrests cells in the G1 phase of the cell cycle. Furthermore, we identify COP1 as a p53-inducible gene, and show that the depletion of COP1 and MDM2 by siRNA cooperatively sensitizes U2-OS cells to ionizing-radiation-induced cell death. Overall, these results indicate that COP1 is a critical negative regulator of p53 and represents a new pathway for maintaining p53 at low levels in unstressed cells.

Arabidopsis thaliana COP1 controls seedling development by negatively regulating light-mediated gene expression⁴. Microarray analysis indicates that *Arabidopsis* COP1 regulates most, if not all, genes that are light responsive^{5,6}. This is exemplified by loss-of-function mutants of the COP/DET/FUS proteins, which have a phenotype that is representative of plants grown in light conditions, even when grown in darkness⁷. Mechanistically, this observation has been attributed to the ability of COP1 to repress positive regulators of light-mediated development, such as LAF1 (ref. 1) and HY5 (ref. 2). Such evidence points to the fact that COP1 is a crucial switch in light-mediated development in plants. However, its role in mammalian cells is less well established³, given that mammalian cells do not undergo photomorphogenesis.

To provide insight into this enigma, we examined COP1 expression in mammalian cells and tissues using real-time polymerase chain reaction (PCR) and *in situ* hybridization techniques. Consistent with previous reports^{3,8}, COP1 was ubiquitously

expressed (data not shown). To gain a further understanding of how COP1 modulates cellular processes, lysates from U2-OS cells stably expressing Flag-tagged COP1 or empty vector were immunoprecipitated with anti-Flag, and bound proteins were eluted via Flag peptide and analysed by SDS–polyacrylamide gel electrophoresis (PAGE). Intriguingly, mass spectrometry analysis revealed five matching peptides from the tumour-suppressor protein p53 (Fig. 1a). To confirm that this interaction is *bona fide*, p53-null Saos-2 cells were transfected with p53 and Myc–COP1, immunoprecipitated with anti-p53 (DO-1), and immunoblotted with anti-Myc (Fig. 1b). COP1 was only immunoprecipitated in the presence of transfected p53, indicating that COP1 can interact with p53 in cells. This interaction was also seen with transfected COP1 and endogenous p53 (Fig. 1c). More importantly, there was an interaction between p53 and COP1 at the endogenous protein level in U2-OS cells (Fig. 1d). And *in vitro*-translated haemagglutinin (HA)–COP1 was able to interact with glutathione S-transferase (GST)–p53 *in vitro* (Fig. 1e).

p53 is a potent tumour-suppressor protein^{9,10} that is negatively regulated or mutated in some, if not all, cancers. Because other RING-finger-containing proteins, such as MDM2 (refs 11–14) and Pirh2 (ref. 15), are interacting proteins and negative regulators of p53, we investigated the possibility that COP1 might also negatively regulate p53. The potential for COP1 to modulate the steady-state levels of p53 protein was assessed in Saos-2 cells by transfecting a constant amount of p53 with increasing amounts of Flag–COP1 or a mutant of COP1 that lacks the RING-finger domain, Flag–COP1 Δ RING. Interestingly, transfection of COP1 resulted in a reduction in the steady-state levels of exogenous p53 protein,

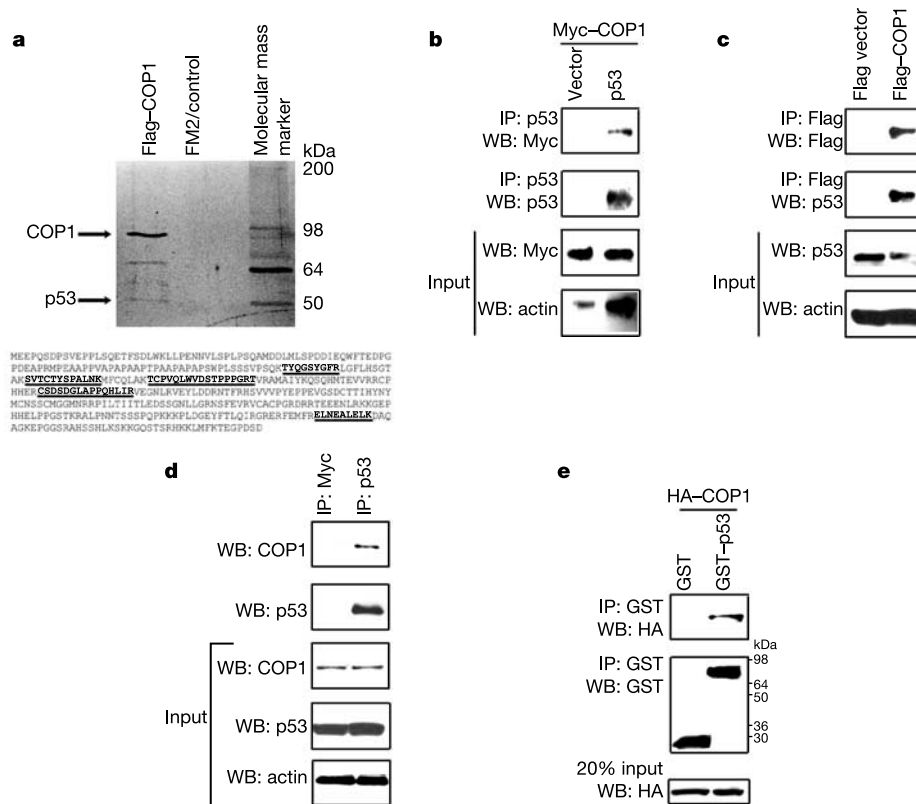


Figure 1 COP1 and its interaction with p53. **a**, Silver-stained SDS gel from Flag-peptide-eluted material. Sequence of the p53 protein (lower panel) with matching peptides by mass spectrometry underlined. **b**, COP1 interacts with exogenous p53 *in vivo*. Saos-2 cells were transfected, immunoprecipitated and immunoblotted (western blot, WB) as indicated. **c**, COP1 interacts with endogenous p53 *in vivo*. U2-OS cells were transfected with or without COP1, immunoprecipitated and immunoblotted as indicated.

d, Endogenous p53 interacts with endogenous COP1. U2-OS cells were immunoprecipitated with anti-p53 or anti-Myc, and immunoblotted with anti-COP1. **e**, COP1 interacts with p53 *in vitro*. GST or GST–p53 was incubated with *in vitro*-translated HA–COP1, and bound COP1 was detected by anti-HA immunoblotting. The lower panel represents 20% total input of HA–COP1.

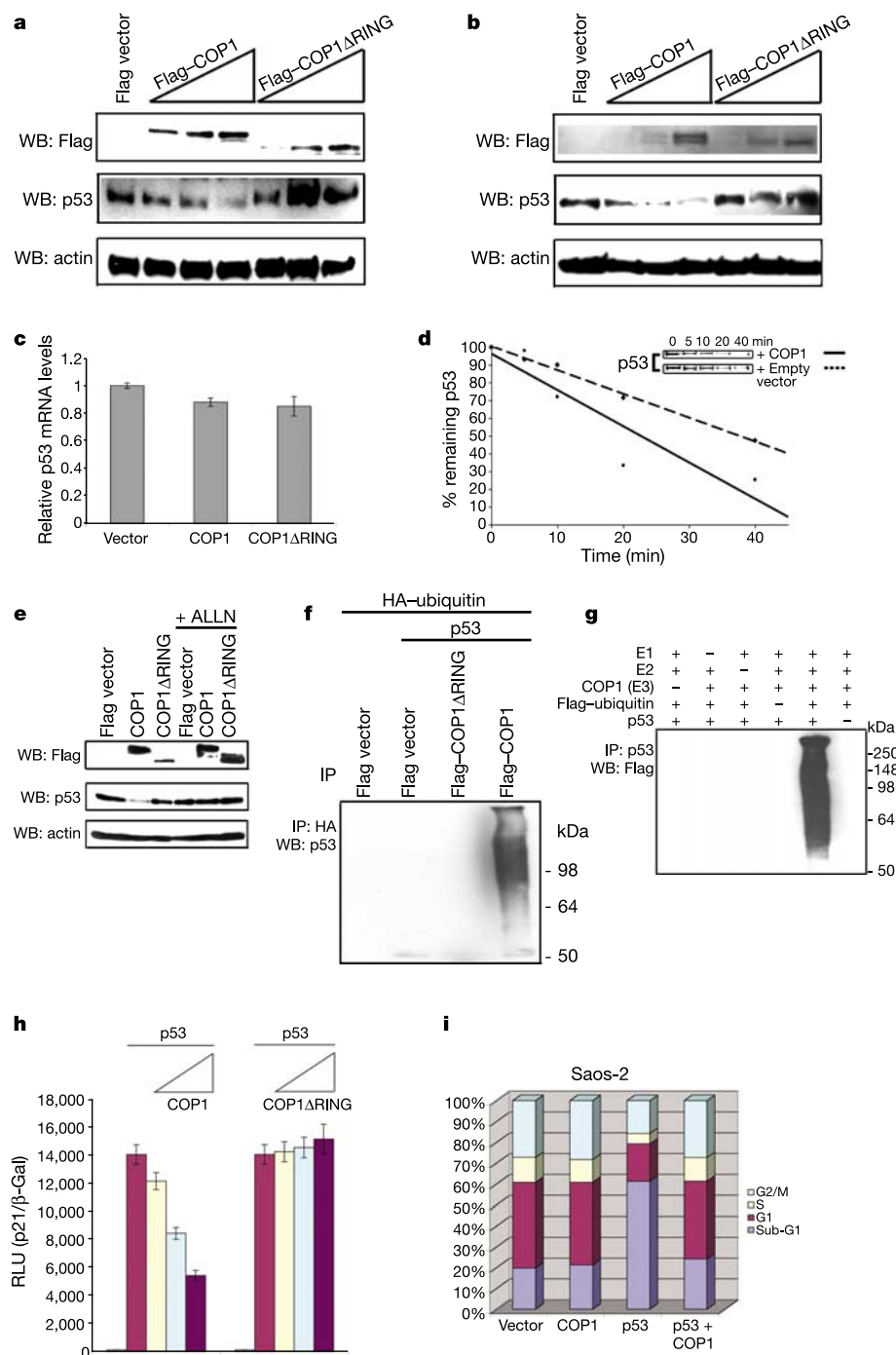


Figure 2 COP1 negative regulation of p53. **a**, COP1 hinders the steady-state level of exogenous p53 protein. COP1 or COP1ΔRING was transfected with p53 into Saos-2 (p53-null) cells, and steady-state levels of p53 were assessed by immunoblotting. **b**, COP1 hinders steady-state levels of endogenous p53 protein. COP1 or COP1ΔRING was transfected into U2-OS cells and steady-state levels of endogenous p53 were assessed by immunoblotting. **c**, COP1 does not alter p53 mRNA levels. RNA from U2-OS cells transfected with COP1 or COP1ΔRING was extracted, and p53 mRNA levels were assessed by real-time PCR and normalized to β-actin mRNA. **d**, COP1 increases p53 turnover. HEK293T cells were transfected with or without COP1, and the half-life of endogenous p53 was determined by pulse-chase metabolic labelling. **e**, COP1 degradation of p53 requires a functional 26S proteasome. U2-OS cells were transfected as in **b**, except that cells were treated with or without the proteasome inhibitor ALLN for 6 h before collecting lysates. **f**, COP1 promotes ubiquitination of p53 *in vivo*. Saos-2 cells

were transfected with HA-ubiquitin, p53 and COP1 or COP1ΔRING, and treated with ALLN. Ubiquitinated p53 was detected by immunoprecipitation with anti-HA and immunoblotting with anti-p53. **g**, COP1 directly ubiquitinates p53 *in vitro*. Bacterially expressed and purified ubiquitin components were incubated with *in vitro*-translated and immunoprecipitated p53 as indicated. Ubiquitinated p53 was detected with anti-Flag and is represented as products >53 kDa. **h**, COP1 dampens the transactivation function of p53 on the p21 promoter. Saos-2 cells were transfected with p53 and COP1 or COP1ΔRING with a p21-Luc reporter and internal control pCMV-β-Gal plasmid. RLU, relative light units. **i**, COP1 prohibits p53-dependent apoptosis. Saos-2 cells were transfected with p53 and COP1 as indicated. Transiently transfected cells were selected by co-transfection of EGFP, and the resultant sub-G1 population was assessed by propidium iodide staining for cell-death estimation.

which was abrogated when the RING-finger domain on COP1 was deleted (Fig. 2a). Furthermore, COP1 was able to decrease the steady-state levels of endogenous p53 in U2-OS cells (Fig. 2b). To gain further insight into the mechanism of this reduction in p53 protein levels, real-time PCR was carried out on the *p53* gene to determine whether COP1 inhibited *p53* gene transcription (Fig. 2c). Transfection of Flag-COP1 or Flag-COP1 Δ RING in U2-OS cells resulted in no significant change in p53 messenger RNA levels. However, using pulse-chase analysis, transfection of COP1 in human embryonic kidney HEK293T cells showed a clear reduction in the half-life of p53 relative to the empty vector control, indicating that COP1 actively increases the turnover of p53. Given that the half-life of p53 is considerably shortened by COP1 overexpression, and that this is independent of *p53* mRNA effects or any direct effect on MDM2 protein levels (data not shown), it is probable that this effect may be post-translational. To test this hypothesis, U2-OS cells were transfected with Flag-COP1 or Flag-COP1 Δ RING and treated with the proteasome inhibitor ALLN (Fig. 2e). The addition of ALLN markedly increased the level of p53 protein, which was previously reduced by transfection of Flag-COP1, indicating that COP1 directs p53 for proteasome-mediated degradation. To determine whether COP1-mediated degradation of p53 by the proteasome is a consequence of p53 ubiquitination, Saos-2 cells were transfected with p53, Flag-COP1 or Flag-COP1 Δ RING, and HA-ubiquitin (Fig. 2f). Immunoprecipitation with anti-HA and immunoblotting with anti-p53 revealed the presence of ubiquitinated species of p53 only upon co-transfection of Flag-COP1, suggesting that COP1 targets p53 for degradation via the proteasome by ubiquitination.

To ascertain whether COP1 acts through MDM2 or Pirh2 to modulate p53 degradation, transient transfections with p53 and COP1 or COP1 Δ RING were carried out in p53^{-/-}/MDM2^{-/-} mouse embryo fibroblasts (MEFs), or in Saos-2 cells depleted of Pirh2 by siRNA, and their ability to affect p53 steady-state levels was assessed (Supplementary Fig. 1a, b). Essentially, COP1 was able to recapitulate the negative regulation of p53 achieved in MDM2- and Pirh2-containing cells.

COP1 has inherent E3 ligase activity^{1,8} and can use LAF1 as a

substrate *in vitro*. In our studies, COP1 modulated the ubiquitination of p53 *in vivo* (Fig. 2f), so we sought to determine whether p53 is a direct substrate for COP1 *in vitro*. GST-COP1 was expressed in *Escherichia coli*, purified, and subsequently used for ubiquitination assays with *in vitro*-translated p53 (Fig. 2g). Only with all of the reaction components and p53 present was COP1 able to directly ubiquitinate p53. These data indicate that COP1 serves as an E3 ligase for p53, identifying p53 as the first direct mammalian substrate for COP1.

To determine the effect of COP1 overexpression on p53-dependent transactivation, Saos-2 cells were transfected with p53 and COP1 or COP1 Δ RING, and *p21*-luciferase (Luc) or *bax*-Luc (Fig. 2h and Supplementary Fig. 2a, respectively). The addition of COP1 markedly reduced the ability of p53 to transactivate from the *p21* and *bax* promoter. Moreover, the transactivation ability of endogenous p53, as well as that induced by DNA damage, was in the same manner abrogated by the overexpression of COP1 (Supplementary Fig. 2b).

To assess further the ability of COP1 to negatively regulate p53, we ascertained whether COP1 could inhibit p53-induced cell death in Saos-2 cells (Fig. 2i). Transfection of p53 markedly increased the population of cells in the SubG1 population; nevertheless, this was strikingly prohibited by the co-transfection of COP1, suggesting that COP1 inhibits p53-induced cell death.

To uncover the role of COP1 in a more physiological setting, endogenous COP1 was subject to ablation by siRNA, and any effect on endogenous p53 steady-state levels was assessed (Fig. 3a). Depletion of COP1 by siRNA in U2-OS cells caused a pronounced accumulation of p53 protein. These results were reproducible with two further independent siRNA oligonucleotides (Supplementary Fig. 3), indicating that COP1 negatively regulates p53 in unstressed cells. With such a dramatic accumulation of p53 protein levels, we next ascertained whether p53 could induce its downstream target genes, *p21* and *Pirh2*. Using real-time PCR, we observed a pronounced increase in *p21* and *Pirh2* mRNA in response to COP1 ablation, indicating an increase in p53-dependent transcription. Furthermore, total p21 protein levels were dramatically increased by the introduction of COP1 siRNA. In contrast, the transfection of

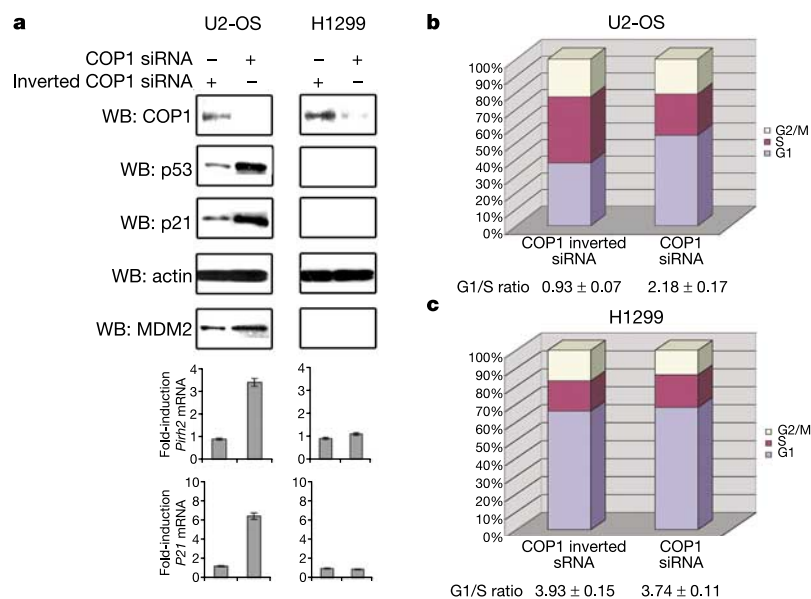


Figure 3 siRNA ablation of COP1 causes an accumulation of p53 protein and induces a G1 arrest. **a**, siRNA ablation of COP1 increases the steady-state level of p53 protein and increases transactivation of the *p21* and *Pirh2* genes. U2-OS and H1299 cells were transfected with siRNA oligonucleotides to COP1 or COP1 inverted as a control. Lysates were immunoblotted with the indicated antibodies and mRNA of target genes was

detected by real-time PCR. **b, c**, siRNA ablation of COP1 induces a G1 arrest that is p53-dependent. U2-OS (**b**) or H1299 (**c**) cells were transfected with COP1 or inverted COP1 siRNA oligonucleotides, and cell-cycle profile was determined by propidium iodide staining and FACS.

COP1 siRNA oligonucleotides into the p53-null cell line H1299 had no effect on p21 protein or on *p21* and *Pirh2* mRNA levels. Furthermore, depletion of COP1 in U2-OS cells caused a prominent increase in the G1/S ratio, but failed to have any effect in the H1299 cells (Fig. 3c), indicating that a G1 arrest occurred in a p53-dependent manner.

To confirm that COP1 mediates p53 turnover in unstressed cells, pulse-chase analysis was carried out in U2-OS cells depleted of COP1 (Fig. 4a). Ablation of COP1 protein increased the half-life of p53 2-fold relative to control. Pirh2 and MDM2 were also ablated by siRNA for pulse-chase analysis and comparison with COP1. Depletion of Pirh2 increased the half-life of p53 1.5-fold, whereas MDM2 ablation increased the half-life of p53 2-fold over control. To place COP1 in the context of MDM2 and Pirh2, co-ablation of COP1 and Pirh2 by siRNA resulted in a further enhancement of p53 half-life, 3.5-fold over control. Surprisingly, ablation of COP1 and

MDM2 together resulted in a synergistic 8-fold increase in p53 half-life. This was not further enhanced by simultaneous Pirh2 depletion, suggesting that the half-life of p53 had reached a maximum in this particular system. Reporter assays were carried out to measure the transactivation function of p53 (Fig. 4b). Consistent with an increase in the half-life of p53, there was a modest increase in transactivation from the *p21* promoter upon ablation of COP1, Pirh2 or MDM2, with the highest stimulation derived from ablation of MDM2. These observations were also confirmed at the protein level (Fig. 4c). This agrees with the ability of MDM2 not only to increase p53 turnover, but also to repress the transactivation activity of p53 (refs 16, 17). There was also a pronounced stimulation from the *p21* promoter and an increase in protein levels in response to co-ablation of COP1 and MDM2 (Fig. 4a, b). It is noteworthy that in response to ablation of all E3 ligases, there was no further enhancement of p53 or p21 protein

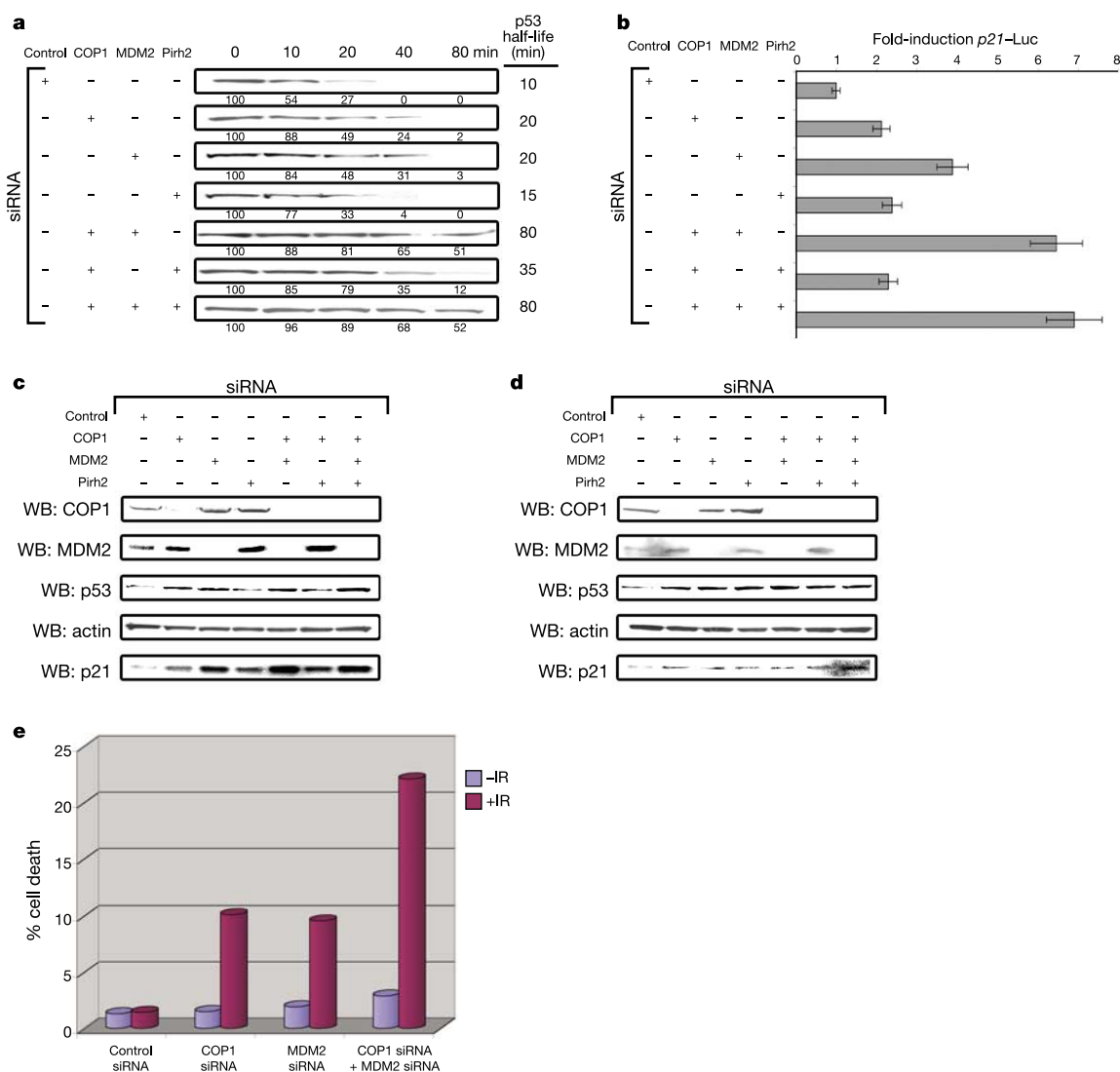


Figure 4 COP1, Pirh2 and MDM2 ablation by siRNA stabilizes and activates p53. **a**, Ablation of COP1 stabilizes p53. U2-OS cells were transfected with siRNA oligonucleotides to COP1, MDM2 and/or Pirh2, and subsequently pulse-chased for the indicated period of time. Lysates were then harvested and immunoprecipitated with anti-p53 and quantified on a phosphorimager. **b**, The transactivation activity of p53 is increased upon diminishment of COP1. U2-OS cells were transfected with the indicated siRNA oligonucleotides and *p21*-Luc reporter, and relative transactivation activity was determined by normalizing luciferase to an internal transfection control, pCMV- β -Gal activity. **c**, p53 and the downstream target p21 accumulate at the protein level in

response to ablation of COP1 and are further stimulated by co-ablation of MDM2. Lysates from **b** were probed with antibodies to p53, p21, COP1 and MDM2. **d**, COP1 negatively regulates p53 in normal cells. BJ fibroblasts were transfected with siRNA oligonucleotides as in **c**, and lysates were harvested and immunoblotted with antibodies to p53, p21, actin, MDM2 and COP1. **e**, Ablation of COP1 and MDM2 by siRNA sensitizes U2-OS cells to IR-induced cell death. U2-OS cells were pre-treated with siRNA oligonucleotides as indicated, and subjected to 20 Gy IR and cell death determined by propidium iodide staining.

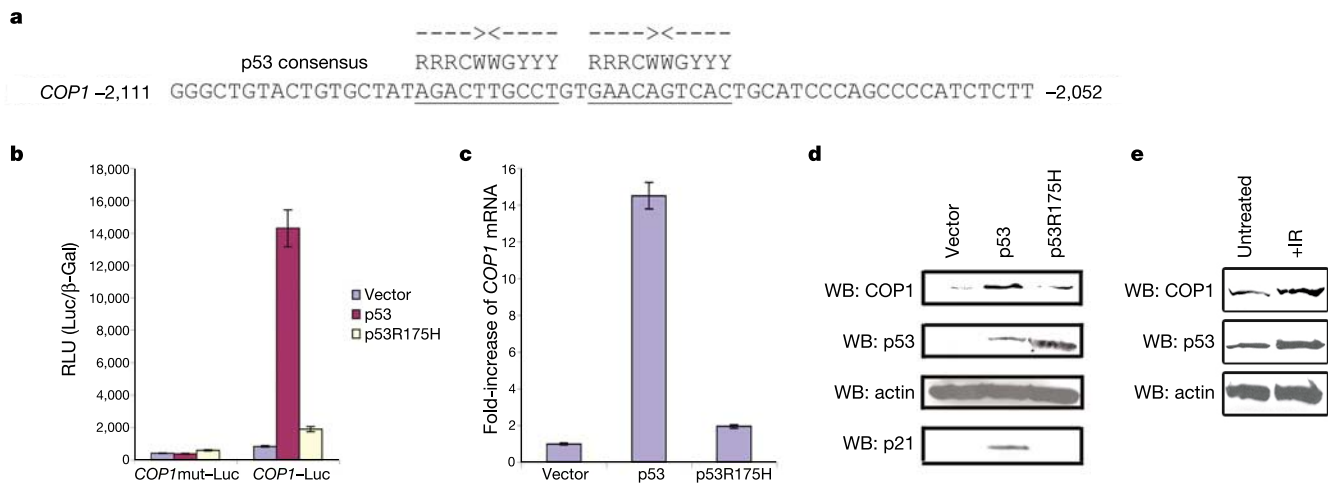


Figure 5 *COP1* is a p53-inducible gene. **a**, Identification of a p53 consensus site on the *COP1* promoter. Underlined sequence represents a p53 decamer. **b**, The p53 consensus site from *COP1* promoter is able to stage p53-dependent transcription. Two copies of the *COP1* consensus site for p53 were introduced upstream of the pGL3-Promoter luciferase reporter construct and co-transfected with wild-type p53 or mutant p53 R175H. Relative activity was determined by normalizing to pCMV- β -Gal activity. **c**, *COP1*

mRNA is induced by p53. H1299 cells were transfected with p53 or R175H mutant, and RNA was isolated and subjected to real-time PCR using a probe specific for *COP1* mRNA and normalized to β -actin mRNA. **d**, *COP1* protein levels are increased by p53. Lysates from **c** were immunoblotted with antibodies to p53, p21, actin and *COP1*. **e**, *COP1* protein levels are increased in response to IR. U2-OS cells were irradiated with 10 Gy IR and harvested after 8 h. Lysates were immunoblotted with antibodies to p53, actin and *COP1*.

levels (Fig. 4c), or of promoter activity (Fig. 4b), suggesting that proliferating cells may have a limited threshold for p21 and/or p53. Furthermore, ablation of *COP1*, Pirh2 or MDM2 resulted in an increase in p53 and p21 steady-state levels in normal BJ fibroblasts, indicating that each ligase has a role in negatively regulating p53 in normal cells (Fig. 4d).

With the observations that ablation of *COP1* and MDM2 resulted in a p53 response, we tested whether depletion of *COP1* and/or MDM2 would restore a normal DNA-damage response in U2-OS cells, because they are known to be highly resistant to ionizing radiation (IR)-induced cell death¹⁸. U2-OS cells were transfected with siRNA oligonucleotides to *COP1* and/or MDM2, and subsequently treated with 20 Gy IR, and cell death was assessed by propidium iodide staining (Fig. 4e). As expected, U2-OS cells pre-treated with control siRNA were highly resistant to cell death after treatment with IR; however, cells pre-treated with *COP1* or MDM2 siRNA resulted in a subpopulation of cells that were sensitive to IR-induced cell death. More strikingly, simultaneous ablation of *COP1* and MDM2 by siRNA resulted in a cooperative sensitization to cell death by IR.

Given that MDM2 and Pirh2 form an autoregulatory feedback loop, we investigated the possibility that *COP1* may also be part of such a regulatory mechanism. Scanning the promoter region of the *COP1* gene revealed the presence of a p53 consensus binding site¹⁹ from -2,094 to -2,073 relative to the predicted transcriptional start site (+1) (Fig. 5a). Two copies of this consensus site (*COP1*-Luc) or mutants of this consensus site (*COP1*mut-Luc) were inserted upstream of a luciferase reporter gene containing a minimal promoter, and transfected into H1299 cells with pcDNA3.1+, p53 or the DNA-binding mutant p53R175H (Fig. 5b). Transfection of p53 increased the luciferase activity of *COP1*-Luc, but not that of the *COP1*mut-Luc reporter construct, whereas the p53R175H mutant failed to have a pronounced effect on *COP1*-Luc or *COP1*mut-Luc. Furthermore, transfection of p53 substantially increased *COP1* mRNA levels within H1299 cells, as assessed by real-time PCR (Fig. 5c). An increase in total *COP1* protein was also detected by western blot with anti-*COP1* (Fig. 5d). In addition, U2-OS cells treated with IR, which activates p53-dependent transcription and stabilizes p53, resulted in an increase in *COP1* protein levels (Fig. 5e) relative to untreated

cells. As controls, p21 and p53 protein levels were also assessed by immunoblotting to verify that the IR insult elicited a p53 response. Taken together, these data suggest that *COP1* is a p53-inducible gene that participates in a negative feedback loop.

Overall, our results provide an insight into the cellular function of mammalian *COP1*. Although there is conservation between *Arabidopsis* *COP1* and human *COP1* in negatively regulating the mammalian homologue of HY5, c-jun⁸, it is likely that the role of *COP1* will be more complex in mammalian cells, given that they do not undergo photomorphogenesis and that plants do not develop cancer²⁰ or have p53 orthologues. The ability of *COP1* to negatively regulate p53 by direct ubiquitination, and c-jun family members by transcriptional inhibition⁸, places *COP1* in the context of a master switch for the cell to undergo growth arrest or apoptosis versus proliferation and growth. Identifying the components of this signalling pathway will undoubtedly provide an opportunity for therapeutic intervention by revealing novel targets for drug design that will potentially maximize the p53 response and sensitize tumour cells to chemotherapeutic agents. □

Methods

Expression vectors, recombinant proteins and antibodies

Flag-*COP1* has been described previously²¹. HA-*COP1* was generated by PCR subcloning *COP1* into pcDNA3.1+ (Invitrogen), and GST-*COP1* was generated by subcloning *COP1* into pGEX6P1 (Pharmacia). pcDNA-Pirh2 was a gift from S. Benchimol. pcDNA3.1+p53, pG13-Luc, p21-Luc, bax-Luc, NS-Luc and pCMV-MDM2 have been described previously²²⁻²⁴. *COP1*-Luc and *COP1*mut-Luc were generated by ligation of oligonucleotides containing two copies of the p53 consensus site, or mutants of the consensus site, derived from the *COP1* promoter into pGL3-Promoter (Promega). All GST recombinant proteins were expressed in *E. coli* strain BL21(DE3) codon + (Stratagene) and subsequently purified using the Glutathione Sepharose 4B batch method (Pharmacia). *In vitro* transcription/translation of recombinant protein was carried out using the T7/T3-coupled TnT kit (Promega). Anti-p53 (DO-1; Calbiochem), anti-p53 (1801; BD Pharmingen), anti-p53 (FL-393; Santa Cruz Biotechnology), anti-p21 (Ab-1; Calbiochem), anti-MDM2 (2A10; Calbiochem), anti-Flag (M2; Sigma), anti-Myc (9E10; Roche), anti-actin (ICN), anti-GST (B-14; Santa Cruz Biotechnology) and anti-HA (Roche) were used according to manufacturer's recommendations. Anti-*COP1* is a mouse monoclonal antibody raised against amino acids 71-270 of human *COP1*.

Cells, transfections, reporter and apoptosis assays

U2-OS, Saos-2, HEK293T and BJ cells were purchased from the American Type Culture Collection (ATCC) and maintained in McCoy's 5A (Invitrogen) or DMEM (Sigma) media. p53^{-/-}/MDM2^{-/-} MEFs were a gift from G. Lozano and were maintained in

DMEM. H1299 cells were a gift from J. Chen and were grown in RPMI. All transfections were carried out using Lipofectamine 2000 (Invitrogen), Oligofectamine (Invitrogen) or Geneporter 2 (Gene Therapy Systems) according to manufacturer's recommendations. To assess the effect of COP1 on steady-state levels of p53, Saos-2 cells were transfected with increasing amounts of Flag-COP1 or Flag-COP1ΔRING with 250 ng pcDNA3.1+p53, and U2-OS cells were transfected with or without increasing amounts (0.5, 1 and 2 μg) of pCMV-Flag-COP1 or pCMV-Flag-COP1ΔRING and treated with 50 μM ALLN for 6 h before cell collection where indicated. For reporter assays, Saos-2 or H1299 cells were transiently transfected with 150 ng of pcDNA3.1+p53 or pcDNA3.1+p53R175H, 100 ng of p21-Luc, bax-Luc, COP1-Luc, COP1mut-Luc or NS-Luc, and 10 ng of pCMV-β-Gal, with or without increasing amounts (0.5, 1 and 2 μg) of pCMV-Flag-COP1 or pCMV-Flag-COP1ΔRING. Luciferase assays were carried out according to manufacturer's instructions (Promega). For p53-induced cell-death assays, Saos-2 cells were transiently transfected with 1 μg of enhanced green fluorescent protein (EGFP) and 5 μg of pcDNA3.1+, pcDNA3.1+p53, pCMV-Flag-COP1 or pcDNA3.1+p53, and 15 μg of pCMV-Flag-COP1 for 48 h. Cells were harvested and stained with propidium iodide for analysis by fluorescence-activated cell sorting (FACS). COP1 siRNA1 (AACUGACCA GAUAACCUUGA), COP1 siRNA1 inverted (AAAGUCCAAUAGAACCAGUC), COP1 siRNA2 (AAGACUUGGAGCAGUGUUAU), COP1 siRNA3 (AAGAGGUGUUGGAG UGUUGAC), Pirh2 siRNA1 (AACTGTGGAATTTGTAGG), Pirh2 B inverted (AAGGAUGUUUAAGGUGUCAA), Pirh2 siRNA2 (AAUGUAAACUUAUGCCUAGCUA), Pirh2 siRNA2 inverted (AAAUCCGAUCCGUUAUCAAUGU) and MDM2 (AAGGAUUUAGACAACCUGAA) siRNA oligonucleotides with 3' dTdT overhangs were synthesized by Genentech or Dharmacon. Control siRNA in experiments refers to a mixture of inverted siRNA oligonucleotides. U2-OS, H1299, Saos-2 and BJ cells were transfected with siRNA oligonucleotides three times at 24–36 h intervals and expanded as necessary to prevent contact inhibition.

Immunoprecipitation, GST pull-down assays and pulse-chase analysis

Cells were lysed in immunoprecipitation (IP) lysis buffer (1% Triton X-100, 150 mM NaCl, 50 mM Tris, pH 7.4, and protease inhibitor mix) or radioimmunoprecipitation assay (RIPA) buffer (0.1% SDS, 1% NP-40, 150 mM NaCl, 0.5% deoxycholate, 50 mM Tris, pH 7.4, and protease inhibitor mix), pre-cleared and immunoprecipitated with target antibody and protein A/G PLUS beads. Identification of COP1-interacting proteins was carried out as previously described²¹, except that U2-OS cells stably expressing Flag-COP1 were generated. GST pull-down assays were carried out with GST or GST-p53 combined with *in vitro* translated HA-COP1 in PBST (PBS with 0.1% Tween 20) and incubated on ice for 1 h. GST-bound proteins were subject to SDS-PAGE and immunoblot with anti-HA and anti-GST. IPs were washed in lysis buffer with high salt as required. Pulse-chase experiments were carried out as previously described¹⁵, except that HEK293T cells were transfected with pCMV-Flag6a or pCMV-Flag-COP1 for 24 h, and U2-OS cells were transfected with siRNA oligonucleotides as indicated.

In vitro ubiquitination assays

For *in vitro* ubiquitination reactions, *in vitro*-translated p53 was immunoprecipitated with anti-p53 (DO-1 and FL-393) and washed five times with IP lysis buffer, and reactions were carried out on protein A/G beads. Ten micrograms of Flag-ubiquitin (Sigma), 20 ng of UbCH5b (A.G. Scientific), 20 ng of rabbit E1 (Sigma) and 500 ng of GST-COP1 (E3), which was pre-incubated with 20 μM ZnCl₂ for 30 min at room temperature, were incubated in a buffer containing 50 mM Tris, pH 7.5, 2 mM ATP, 5 mM MgCl₂, 20 μM ZnCl₂ and 2 mM DTT. After incubation for 2 h at 30 °C with gentle agitation, reactions were boiled in PBST with 1% SDS for 5 min and reduced to 0.1% SDS with PBST for re-immunoprecipitation with anti-p53 (DO-1 and FL-393). Finally, samples were subjected to SDS-PAGE followed by immunoblotting with anti-Flag-HRP (M2) to detect ubiquitinated species of p53.

Received 23 December 2003; accepted 29 March 2004; doi:10.1038/nature02514.
Published online 21 April 2004.

- Seo, H. S. *et al.* LAF1 ubiquitination by COP1 controls photomorphogenesis and is stimulated by SPA1. *Nature* **424**, 995–999 (2003).
- Ang, L. H. *et al.* Molecular interaction between COP1 and HY5 defines a regulatory switch for light control of *Arabidopsis* development. *Mol. Cell* **13**, 213–222 (1998).
- Yi, C., Wang, H., Wei, N. & Deng, X. W. An initial biochemical and cell biological characterization of the mammalian homologue of a central plant developmental switch, COP1. *BMC Cell Biol.* **3**, 30 (2002).
- Neff, M. M., Fankhauser, C. & Chory, J. Light: an indicator of time and place. *Genes Dev.* **14**, 257–271 (2000).
- Ma, L. *et al.* Light control of *Arabidopsis* development entails coordinated regulation of genome expression and cellular pathways. *Plant Cell* **13**, 2589–2607 (2001).
- Ma, L. *et al.* Genomic evidence for COP1 as a repressor of light-regulated gene expression and development in *Arabidopsis*. *Plant Cell* **14**, 2383–2398 (2002).
- Hardtke, C. S. & Deng, X. W. The cell biology of the COP/DET/FUS proteins. Regulating proteolysis in photomorphogenesis and beyond? *Plant Physiol.* **124**, 1548–1557 (2000).
- Bianchi, E. *et al.* Characterization of human constitutive photomorphogenesis protein 1, a RING finger ubiquitin ligase that interacts with Jun transcription factors and modulates their transcriptional activity. *J. Biol. Chem.* **278**, 19682–19690 (2003).
- Oren, M. Decision making by p53: life, death and cancer. *Cell Death Differ.* **10**, 431–442 (2003).
- Jin, S. & Levine, A. J. The p53 functional circuit. *J. Cell Sci.* **114**, 4139–4140 (2001).
- Shimizu, H. & Hupp, T. R. Intracellular regulation of MDM2. *Trends Biochem. Sci.* **28**, 346–349 (2003).
- Honda, R., Tanaka, H. & Yasuda, H. Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Lett.* **420**, 25–27 (1997).
- Kubbutat, M. H., Jones, S. N. & Vousden, K. H. Regulation of p53 stability by Mdm2. *Nature* **387**, 299–303 (1997).

- Haupt, Y., Maya, R., Kazan, A. & Oren, M. Mdm2 promotes the rapid degradation of p53. *Nature* **387**, 296–299 (1997).
- Leng, R. P. *et al.* Pirh2, a p53-induced ubiquitin-protein ligase, promotes p53 degradation. *Cell* **112**, 779–791 (2003).
- Thut, C. J., Goodrich, J. A. & Tjian, R. Repression of p53-mediated transcription by MDM2: a dual mechanism. *Genes Dev.* **11**, 1974–1986 (1997).
- Mirnezami, A. H. *et al.* Hdm2 recruits a hypoxia-sensitive corepressor to negatively regulate p53-dependent transcription. *Curr. Biol.* **13**, 1234–1239 (2003).
- Allan, L. A. & Fried, M. p53-dependent apoptosis or growth arrest induced by different forms of radiation in U2OS cells: p21WAF1/CIP1 repression in UV induced apoptosis. *Oncogene* **18**, 5403–5412 (1999).
- Qian, H., Wang, T., Naumovski, L., Lopez, C. D. & Brachmann, R. K. Groups of p53 target genes involved in specific p53 downstream effects cluster into different classes of DNA binding sites. *Oncogene* **21**, 7901–7911 (2002).
- Doonan, J. & Hunt, T. Cell cycle Why don't plants get cancer? *Nature* **380**, 481–482 (1996).
- Wertz, I. E. *et al.* Human De-etiolated-1 regulates c-Jun by assembling a CUL4A ubiquitin ligase. *Science* **303**, 1371–1374 (2004).
- Dornan, D., Shimizu, H., Perkins, N. D. & Hupp, T. R. DNA-dependent acetylation of p53 by the transcription coactivator p300. *J. Biol. Chem.* **278**, 13431–13441 (2003).
- Dornan, D. & Hupp, T. R. Inhibition of p53-dependent transcription by BOX-1 phospho-peptide mimetics that bind to p300. *EMBO Rep.* **2**, 139–144 (2001).
- Shimizu, H. *et al.* The conformationally flexible S9–S10 linker region in the core domain of p53 contains a novel MDM2 binding site whose mutation increases ubiquitination of p53 *in vivo*. *J. Biol. Chem.* **277**, 28446–28458 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We would like to thank T. Hupp for his contribution of reagents and thought-provoking discussions, G. Lozano for p53^{-/-}/MDM2^{-/-} MEFs, S. Benchimol for the Pirh2 antibody and cDNA, J. Chen for H1299 cells, B. Henzel for mass spectrometry support, C. Reed for the generation of COP1 monoclonal antibodies, M. Vasser and P. Ng for oligonucleotide synthesis and purification, A. Waugh for Bioinformatics support, Genentech Protein Engineering and core DNA sequencing facility for support services, and P. Polakis and members of the Dixit lab for advice and encouragement. I.W. is supported by a PSTP fellowship from the University of California, Davis.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.M.D (dixit@gene.com).

Integrating high-throughput and computational data elucidates bacterial networks

Markus W. Covert*, Eric M. Knight, Jennifer L. Reed, Markus J. Herrgard & Bernhard O. Palsson

Bioengineering Department, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0412, USA

* Present address: Biology Division, California Institute of Technology, 1200 E. California Boulevard, Mail Code 147-75, Pasadena, California 91125, USA

The flood of high-throughput biological data has led to the expectation that computational (or *in silico*) models can be used to direct biological discovery, enabling biologists to reconcile heterogeneous data types, find inconsistencies and systematically generate hypotheses^{1–3}. Such a process is fundamentally iterative, where each iteration involves making model predictions, obtaining experimental data, reconciling the predicted outcomes with experimental ones, and using discrepancies to update the *in silico* model. Here we have reconstructed, on the basis of information derived from literature and databases, the first integrated genome-scale computational model of a transcriptional regulatory and metabolic network. The model accounts for 1,010 genes in *Escherichia coli*, including 104 regulatory genes whose products together with other stimuli regulate the expression of 479 of the 906 genes in the recon-

structured metabolic network. This model is able not only to predict the outcomes of high-throughput growth phenotyping and gene expression experiments, but also to indicate knowledge gaps and identify previously unknown components and interactions in the regulatory and metabolic networks. We find that a systems biology approach that combines genome-scale experi-

mentation and computation can systematically generate hypotheses on the basis of disparate data sources.

We first validated the model, or 'in silico strain' of *E. coli* (iMC1010^{vi}; see ref. 4 for conventions for naming in silico strains), against a data set of 13,750 growth phenotypes⁵ obtained from the ASAP database⁶, and then used this genome-scale model to select

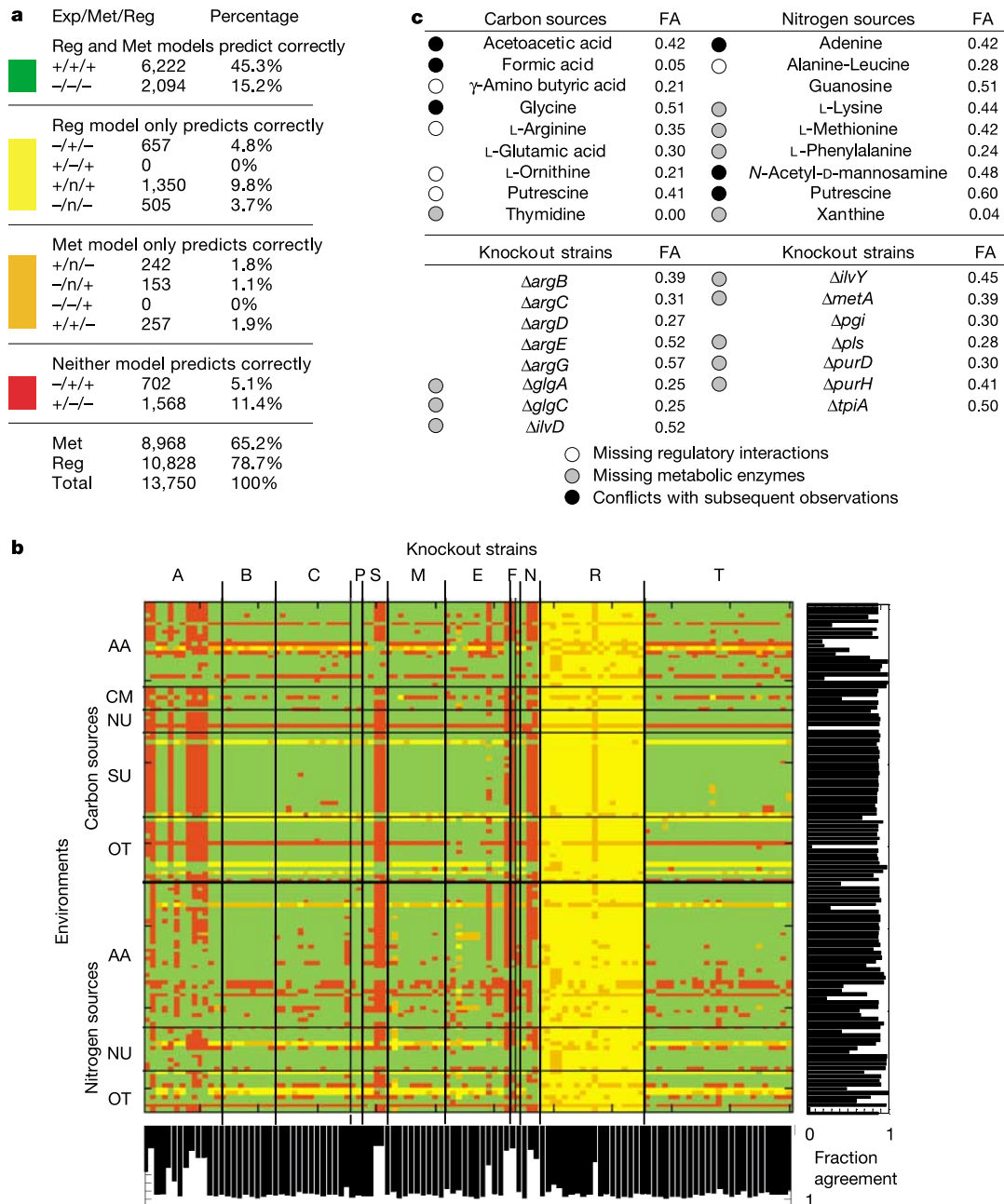


Figure 1 Growth phenotype study. **a**, Comparison of high-throughput phenotyping array data (Exp) with predictions for the *E. coli* network, both considering regulatory constraints (Reg) and ignoring such constraints as a control (Met). Each case is categorized by comparison type (Exp/Met/Reg), and results are listed as '+' (predicted or observed growth), '-' (no growth) or 'n' (for cases involving a regulatory gene knockout not predictable by the Met model). The comparisons are further divided into four subgroups represented by different colours. **b**, Chart showing individual results for each knockout under each environmental condition, with results categorized and coloured as in **a**. The environments involve variation of a carbon or nitrogen source and are further divided into subgroups: AA, amino acid or derivative; CM, central metabolic intermediate; NU, nucleotide or nucleoside; SU, sugar; OT, other. The knockout strains are also divided by

functional group: A, amino acid biosynthesis and metabolism; B, biosynthesis of cofactors, prosthetic groups and carriers; C, carbon compound catabolism; P, cell processes (including adaptation and protection); S, cell structure; M, central intermediary metabolism; E, energy metabolism; F, fatty acid and phospholipid metabolism; N, nucleotide biosynthesis and metabolism; R, regulatory function; T, transport and binding proteins; U, unassigned. Each environment and knockout strain is associated with a fraction of agreement (FA) between regulatory model predictions and observed phenotypes, as shown in the bar charts to the right and below. **c**, Table showing all environments or knockout strains for which FA < 0.60. Of these substrates or knockout strains, 18 point to uncharacterized metabolic or regulatory capabilities in this organism, as indicated (see Supplementary Information for a case-by-case analysis).

transcription factors for prospective gene knockout studies. Comparison with the growth phenotypes showed that experimental and computational outcomes agreed in 10,828 (78.7%) of the cases examined, which is roughly the same success rate achieved in previous studies in *E. coli* and yeast that considered only a few hundred phenotypes^{7–9}. In addition, 2,512 (18.3%) of the cases were predicted correctly only when regulatory effects were incorporated into the metabolic model (see Supplementary Information for details).

The comparisons in this study identified several substrates and knockout strains whose growth behaviour did not match predictions (Fig. 1). Further investigation of these conditions and strains led to the identification of five environmental conditions in which dominant, as yet uncharacterized, regulatory interactions actively contribute to the observed growth phenotype, and five environ-

mental conditions and eight knockout strains that highlight uncharacterized enzymes or non-canonical pathways that are predicted to be used by the organism (Fig. 1; a detailed analysis of the discrepancies is provided in the Supplementary Information).

We wanted to determine the utility of this model-driven approach in elucidating transcriptional regulatory networks. A previous study, which evaluated the consistency between existing gene expression data sets and the known transcriptional regulatory network of *E. coli*, identified the response to oxygen deprivation as a partially consistent module^{10,11}. We therefore targeted this part of the transcriptional regulatory network for further network characterization. Six strains with knockouts of key transcriptional regulators in the oxygen response ($\Delta arcA$, $\Delta appY$, Δfnr , $\Delta oxyR$, $\Delta soxS$ and the double knockout $\Delta arcA\Delta fnr$) were constructed. The messenger RNA expression profiles of these strains, as well as the

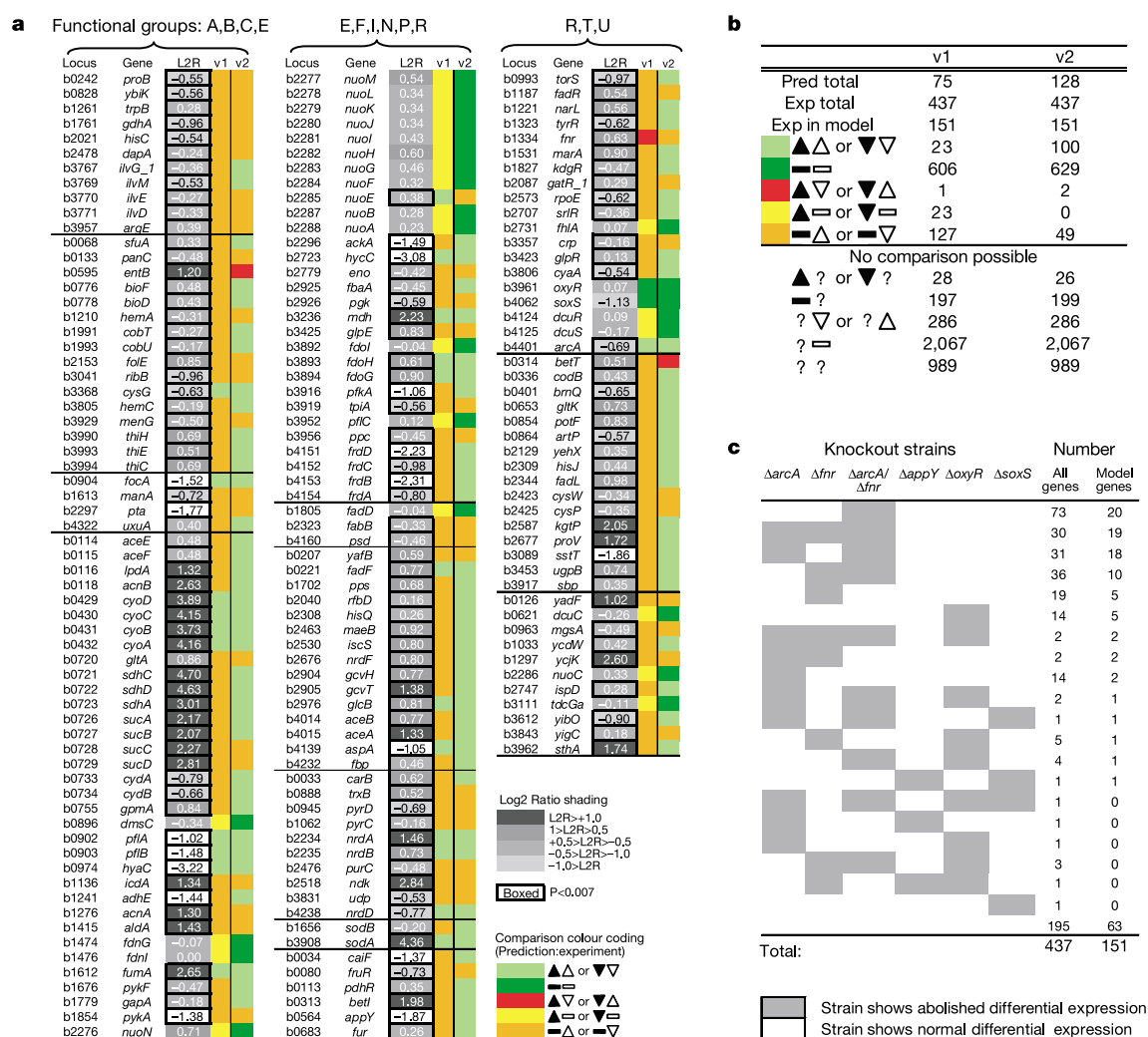


Figure 2 Characterization of the regulatory network related to the aerobic–anaerobic shift. **a**, The locus numbers, gene names and the log₂ ratio (L2R) of gene expression (aerobic to anaerobic) are shown for all model genes with either predicted or observed changes in expression (genes were divided into the same functional groups as in Fig. 1). The L2R values are shaded depending on the magnitude of the expression shift, and those enclosed by a box indicate a statistically significant change in expression ($P < 0.007$, FDR < 5%). Comparisons between the experimental data and model predictions are also shown, where v1 ($iMC1010^{v1}$) and v2 ($iMC1010^{v2}$) designate the model used in the predictions. Filled and open symbols indicate model predictions and experimental data, respectively; rectangles indicate no change in gene expression; triangles indicate a change in expression, as well as the direction of change (upregulated or downregulated).

b, Comparison of the predicted and observed expression changes for the v1 and v2 models. A question mark indicates either that the given gene was not included in the model or that no expression data were obtained for a given shift; other symbols are the same as in **a**. **c**, Systematic perturbation analysis was used to determine the transcription factors responsible for the expression change. The transcription factors knocked out in the six strains are shown on top. Each row indicates a pattern of knockout strains in which differential expression was abolished. The number of genes that show this pattern is indicated on the right. Thus, the first row indicates that for 73 of the 437 genes that showed differential expression in the wild-type strain (or 20 of the 151 genes accounted for by the model), the observed differential expression was abolished only in the $\Delta arcA\Delta fnr$ knockout strain.

wild-type strain, were measured in aerobic and anaerobic glucose minimal medium conditions. The data were analysed¹² in the context of *iMC1010*^{v1} predictions to identify new interactions in the regulatory network (Fig. 2).

Expression profiling of the wild-type strain identified 437 genes that experienced a significant change in transcription in response to oxygen deprivation (*t*-test, multiple testing corrected to give a false discovery rate (FDR) of less than 5%); of these, 151 genes were included in *iMC1010*^{v1}. Computationally, 75 genes were predicted by *iMC1010*^{v1} to show differential expression in response to oxygen deprivation. These 75 genes could be divided into three categories: 23 agreed with measured expression changes; 24 had a predicted expression change that was either not found to be statistically significant in the experimental data (23/24) or in a direction opposite to that of the experimental data (1/24); and for 28 genes there were no expression data available (transcript abundance was determined to be 'absent' for two or more of the replicates). Thus, of the 47 (=23 + 24) differentially expressed genes that could be compared between the model computation and experiment, 23 (or 49% accuracy) agreed. Considering the overall number of genes in the model for which there were experimental data, the overlap (23) between the sets of predicted (47) and experimentally detected (151) differentially expressed genes is significant in comparison to a model that would randomly predict expression changes ($P < 0.005$ on the basis of a cumulative binomial distribution). There were 151 genes that were differentially expressed and included in the model; however, with only 23 (or 15% coverage) correctly computed, there is much room for expanding the transcriptional regulatory network in *iMC1010*^{v1} on the basis of the experimental data (Fig. 3).

To understand which transcription factors are involved in regulating these differentially expressed genes after oxygen deprivation, we compared the gene expression data for the wild-type and each knockout strain separately. Using two-way analysis of variance

(ANOVA), we could determine whether the differential expression was significantly altered in the knockout strain as compared with the wild type. A large portion of the expression changes observed for the wild-type strain were not significantly affected in any of the knockout strains (195/437 or 44.6% of genes overall, 63/151 or 41.7% of genes in the model, FDR < 5%), suggesting that none of the five transcription factors studied here regulates the expression of these genes or that combinatorial interactions between multiple transcription factors are involved in regulation. For the remainder of the genes, differential expression was abolished in one or more of the knockout strains (Fig. 2c).

The ANOVA-based identification of transcription factors that influence differential expression of specific genes enabled us systematically to rewrite, relax or remove various regulatory rules in the model to resolve the discrepancies between *iMC1010*^{v1} and the experimentally determined wild-type differential gene expression. For many (81) of the genes, a regulatory rule already existed and had to be reconciled with our new data to accommodate the newly determined transcription factor dependencies. For genes where none of the knockouts abolished differential expression, we simply based a new regulatory rule on the presence of oxygen rather than a transcription factor (39 genes). By contrast, for genes where a change in expression was predicted but not observed, we removed oxygen dependency from the existing regulatory rule (23 genes). In addition, for 12 genes the predicted expression changes agreed with the observed expression in the wild type, but our knockout perturbation analysis indicated that the transcription factors involved in the regulation differed from previously reported data and the model needed to be changed (all new regulatory rules are detailed in the Supplementary Information).

The updated model (*iMC1010*^{v2}) was used to recalculate all of the predictions for both the aerobic and anaerobic expression data and the high-throughput phenotyping arrays. Note that *iMC1010*^{v2} accounts for the same genes as *iMC1010*^{v1} but has different regulatory interactions among the gene products and oxygen as an environmental variable. We found agreement between model predictions and the gene expression data to be substantially higher using the *iMC1010*^{v2} model, as expected (Fig. 2c). Specifically, 100 of the 151 expression changes were correctly computed with *iMC1010*^{v2}, and the number of false-positive predictions (Fig. 2, yellow boxes) was reduced to zero. In resolving many of the cases of unpredicted differential expression (Fig. 2, orange boxes), we found that implementation of the ANOVA-derived rule resulted in the inability of the wild-type or knockout *in silico* strain to grow aerobically or anaerobically on glucose, or under other conditions where growth had been previously established (for example, wild-type and knockout strain average growth rate under aerobic conditions, 0.68 ± 0.04 per hour; anaerobic, 0.43 ± 0.07 per hour). Such cases may be thought of as an 'overfit' of the microarray data. Accordingly, we relaxed the regulatory rule for these genes (42 in total) to allow for a correct phenotype prediction. Comparisons for the high-throughput phenotyping data revealed very little difference from Fig. 1 (only 11 out of 13,750 cases were affected; see Supplementary Information).

The iterative modification of the regulatory rules led to three main observations. First, some of the results of the knockout perturbation analysis are complex enough to make boolean rule formulation difficult. For example, the interplay of *Fnr* and *ArcA* can lead to complex behaviours where the expression change observed in wild type is abolished in the $\Delta arcA$ or the Δfnr strains, but not in the $\Delta arcA \Delta fnr$ strain. Such complex interplay among transcription factors can lead to specialized expression changes, as observed in the *cydAB* response to anaerobic, microaerobic and aerobic conditions^{13,14}.

Second, in revising regulatory rules for transcription factors we found that whereas in some cases, such as *arcA*, expression of a regulatory protein correlates positively with its activity, in several

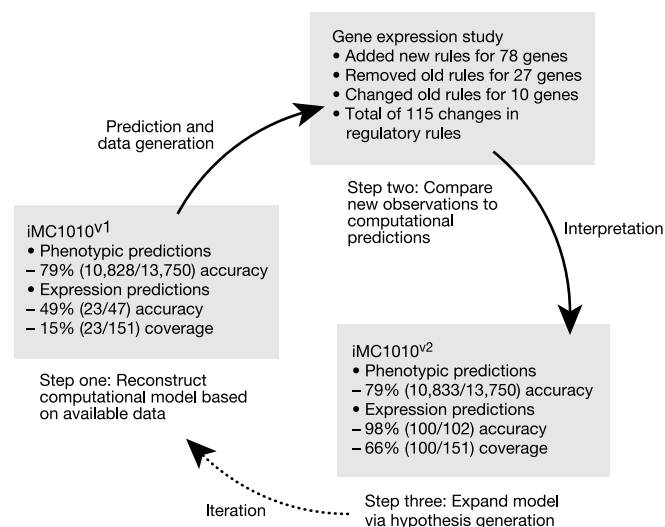


Figure 3 Biological network elucidation by a model-centric approach. Metabolic and regulatory networks may be expanded by using high-throughput phenotyping and gene expression data coupled with the predictions of a computational model. If model predictions are consistent with experimental observations, the network is adequately characterized. If not, the model identifies a knowledge gap and may be used to update, validate and generate hypotheses about organism function. Accuracy refers to the percentage of model predictions that agree with experimental data; coverage indicates the percentage of experimental changes predicted correctly by the model.

cases, including *fnr*, *betI* and *fur* among others, the mRNA level of a regulatory gene is reduced when the protein is in fact activated. For example, under anaerobic conditions when Fnr is known to be active¹¹, its expression is significantly reduced. Such behaviour, underscored by similar observations of mRNA transcript levels and corresponding protein product abundance in yeast¹⁵, suggests that the identification of regulatory networks, and transcription factors involved in regulation in particular, will not be accomplished by the determination of co-regulated gene sets alone.

Third, many of these gene expression changes involve complex interactions and indirect effects. Transcription factors may be affected, for example, by the presence of fermentation by-products or the build up of internal metabolites. Such effects would be extremely difficult to identify or account for without a computational model.

In summary, we find that the reconciliation of high-throughput data sets with genome-scale computational model predictions enables systematic and effective identification of new components and interactions in microbial biological networks. Our study illustrates only the first round of an iterative model building strategy where an initial model based on literature-derived information (*iMC1010*^{v1}) is used to design informative experiments and then updated on the basis of the new experimental data obtained (*iMC1010*^{v2}). Another round of perturbation experiments will lead to *iMC1010*^{v3}, and so on. We expect that after an effort of some years and many iterations of this process, regulatory network elucidation for *E. coli* will be essentially complete. □

Methods

Computational model

We constructed the model of the *E. coli* metabolic and regulatory network by identifying network components, their functions and interactions from the primary literature^{4,9,16}. Many approaches have been developed to analyse large-scale metabolic^{17–22} and transcriptional regulatory^{23–25} networks. Growth and gene expression simulations were done by regulated flux-balance analysis, which combines linear optimization to determine a growth-optimized metabolic flux distribution with logic statements to simulate the effects of regulatory processes over time. The whole model construction and simulation process has been described elsewhere in detail²⁶.

Strains and culture

The parent strain for knockout strains in this study was K-12 MG1655 (ref. 27), and all deletion strains were generated as described²⁸. Growth experiments for the gene expression study were done on M9 glucose medium (2 g l⁻¹) under aerobic and anaerobic conditions, as described¹⁷. The growth data contained in the ASAP database were obtained by using high-throughput phenotype arrays (Biolog)⁵. In some cases (where the viability of a particular environment was unclear from the phenotype array data), we cross-validated the ASAP phenotyping data by culturing the wild-type strain under the given conditions in our laboratory (see Supplementary Information).

Gene expression profiling and analysis

All gene expression measurements were done at least in triplicate. Samples were stabilized by using RNeasy Protect bacterial reagent (Qiagen), and total RNA was isolated from exponentially growing cells using a RNeasy mini kit (Qiagen) in accordance with the manufacturer's protocols (see <http://www1.qiagen.com>). The RNA (10 µg) was then used as the template for complementary DNA synthesis, the product of which was fragmented, labelled and hybridized to an *E. coli* Antisense Genome Array (Affymetrix), which was washed and scanned to obtain an image in accordance with the manufacturer's protocols (see <http://www.affymetrix.com>). The image files were processed and expression values were normalized using dChip software²⁹. We used quantitative real-time polymerase chain reaction with reverse transcription (RT-PCR) to validate expression changes for selected genes. The statistical significance of expression changes for each gene and each strain between aerobic and anaerobic conditions was determined by a *t*-test (log-transformed data, equal variance).

For each deletion strain, we used a two-way ANOVA (strain as the first factor and aerobic or anaerobic condition as the second factor) to determine whether the differential expression observed in the wild-type strain was significantly altered in the deletion strain by determining the statistical significance of the strain-condition interaction effect. For both the *t*-test and the ANOVA analysis, correction for multiple testing was done by using the Benjamini-Hochberg false discovery rate procedure³⁰, which determines the *P*-value cut-off for each test separately by estimating the FDR resulting from using a particular *P*-value cut-off. The false discovery rate refers to the fraction of true null tests out of all the tests called significant and an FDR of 5% was used for all tests. All gene expression data and

the relevant information (such as the MIAME checklist) are provided in the Supplementary Information.

Received 22 November 2003; accepted 1 March 2004; doi:10.1038/nature02456.

- Ideker, T., Galitski, T. & Hood, L. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372 (2001).
- Palsson, B. O. The challenges of *in silico* biology. *Nature Biotechnol.* **18**, 1147–1150 (2000).
- Kitano, H. Systems biology: a brief overview. *Science* **295**, 1662–1664 (2002).
- Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. O. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **4**, R54.1–R54.12 (2003).
- Bochner, B. R. New technologies to assess genotype-phenotype relationships. *Nature Rev. Genet.* **4**, 309–314 (2003).
- Glasner, J. D. *et al.* ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Res.* **31**, 147–151 (2003).
- Forster, J., Famili, L., Palsson, B. O. & Nielsen, J. Large-scale evaluation of *in silico* gene knockouts in *Saccharomyces cerevisiae*. *Omics* **7**, 193–202 (2003).
- Edwards, J. S. & Palsson, B. O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl Acad. Sci. USA* **97**, 5528–5533 (2000).
- Covert, M. W. & Palsson, B. O. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **277**, 28058–28064 (2002).
- Herrgard, M. J., Covert, M. W. & Palsson, B. O. Reconciling gene expression data with known genome-scale regulatory network structures. *Genome Res.* **13**, 2423–2434 (2003).
- Salmon, K. *et al.* Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *J. Biol. Chem.* **278**, 29837–29855 (2003).
- Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 (2001).
- Compan, I. & Touati, D. Anaerobic activation of *arcA* transcription in *Escherichia coli*: roles of Fnr and ArcA. *Mol. Microbiol.* **11**, 955–964 (1994).
- Cotter, P. A., Melville, S. B., Albrecht, J. A. & Gunsalus, R. P. Aerobic regulation of cytochrome *d* oxidase (*cydAB*) operon expression in *Escherichia coli*: roles of Fnr and ArcA in repression and activation. *Mol. Microbiol.* **25**, 605–615 (1997).
- Griffin, T. J. *et al.* Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol. Cell Proteomics* **1**, 323–333 (2002).
- Reed, J. L. & Palsson, B. O. Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. *J. Bacteriol.* **185**, 2692–2699 (2003).
- Edwards, J. S., Ibarra, R. U. & Palsson, B. O. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol.* **19**, 125–130 (2001).
- Price, N. D., Papin, J. A., Schilling, C. H. & Palsson, B. O. Genome-scale microbial *in silico* models: the constraints-based approach. *Trends Biotechnol.* **21**, 162–169 (2003).
- Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99**, 15112–15117 (2002).
- Burgard, A. P. & Maranas, C. D. Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. *Biotechnol. Bioeng.* **74**, 364–375 (2001).
- Burgard, A. P., Vaidyaraman, S. & Maranas, C. D. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnol. Prog.* **17**, 791–797 (2001).
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
- Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet.* **31**, 64–68 (2002).
- Gutierrez-Rios, R. M. *et al.* Regulatory network of *Escherichia coli*: consistency between literature knowledge and microarray profiles. *Genome Res.* **13**, 2435–2443 (2003).
- Bar-Joseph, Z. *et al.* Computational discovery of gene modules and regulatory networks. *Nature Biotechnol.* **21**, 1337–1342 (2003).
- Covert, M. W., Schilling, C. H. & Palsson, B. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* **213**, 73–88 (2001).
- Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
- Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
- Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA* **98**, 31–36 (2001).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Stadslev and A. Fleming for assistance with computation; Z. Zhang and A. Raghunathan for experimental assistance; the Perna and Blattner laboratories for access to the high-throughput phenotyping data in the ASAP database; and the NIH for funding and support. M.W.C. and B.O.P. designed the project and were involved in all phases of the study; E.M.K. carried out experiments; J.L.R. reconstructed the model, ran simulations and did the phenotyping analysis; M.J.H. did the statistical analysis of the gene expression data.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature

Correspondence and requests for materials should be addressed to B.O.P. (palsson@ucsd.edu). The gene expression data are available online in GEO (<http://www.ncbi.nlm.nih.gov/geo/>), accession number GSE1121.

Proteomics in multiplex

Protein arrays and protein assays in parallel are enabling researchers to look at protein interactions and activity on a large scale, as Lisa Melton finds out.

Mapping the protein universe is a daunting task. Charting the interactions of proteins in the human proteome, studying their function and how their expression changes in health and disease, may take decades. Compared with about 40,000 genes in the genome, the human proteome is predicted to top 1 million. "The existence of more than a million protein isoforms in a single human being, in different cells, during his or her lifetime is not unreasonable," speculates Rolf Apweiler, Swiss-Prot coordinator at the European Bioinformatics Institute in Hinxton, UK. "These are all guesses," he cautions, but, so far, over 300 known post-translational modifications have been discovered, and these can occur in different combinations, on different proteolytically processed forms and splice variants. A growing array of tools is allowing scientists to advance the study of the proteome, boost drug discovery and pave the way to personalized medicine.

Traditional proteomics uses techniques such as two-dimensional gel electrophoresis

or chromatography, combined with mass spectrometry, to separate and identify proteins on a large scale (see *Nature* 424, 581–587; 2003). Tracking protein interactions and activity and determining function at this scale is an even greater task, but is beginning to be tackled by protein arrays and multiplexed assays.

To cancer researcher Julio Celis, coordinator of the Danish Centre for Translational Breast Cancer Research in Copenhagen, protein arrays are a breakthrough because they allow many different proteins to be tracked simultaneously. "Discoveries have to be made using very small amounts of clinically relevant sample, and you will want to probe, say, thousands of tissues, both normal and pathological. Multiplexing allows that; it makes life a lot easier," he says.

The commonest type of protein array, or protein chip, has a large number of spots of either proteins or their ligands arranged in a predefined pattern, arrayed by robots onto coated glass slides, microplates or membranes. The array may consist of anti-

bodies to bind proteins of interest (see *Nature* 426, 725–731; 2003), enzymes that will interact with substrates, or substrates or ligands that will interact with applied proteins.

Arrays made easy

One technical hurdle to producing a reliable array is shelf-life. Most protein arrays use antibodies as the immobilized 'probe'. "The process of depositing an antibody onto a solid surface can denature the antibody and affect its recognition properties," says Michael Hadjisavas, strategic marketing manager for Sigma-Aldrich in St Louis, Missouri. "We've developed the know-how to select the most desirable antibodies and maintain their biological activity on an array for up to two years." Sigma's off-the-shelf chips provide a hassle-free introduction to protein arrays. The Panorama Ab microarray kit for cell signalling has 224 different antibodies, arrayed on nitrocellulose-coated glass slides. Proteins extracted from cells or tissues are labelled with the fluorescent dyes Cy3 and Cy5, and the protocol is similar to that of a DNA microarray, with the two labelled protein pools being mixed and applied to the array. The assay takes only 4–5 hours, and can be measured



Julio Celis: keeping track of proteins.

DO-IT-YOURSELF ARRAYS

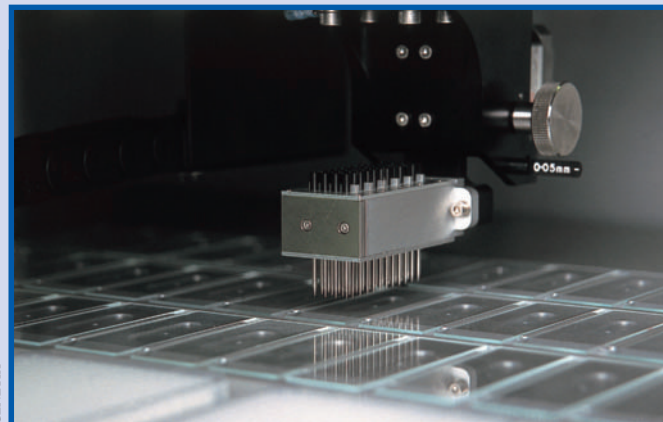
For those who want to spot their own arrays, a number of companies provide standard-sized 25 × 76 mm coated slides especially formulated for protein arrays. Proteins arrayed on the nitrocellulose-coated FASTSlide from Schleicher and Schuell Bioscience in Cologne, Germany, are stable for up to three years at room temperature, according to the company. SuperProtein

slides coated with hydrophobic polymer from TeleChem International in Sunnyvale, California, and the MaxiSorp black polymer-coated slides from Nalge Nunc in Rochester, New York, are also designed for protein arrays. 3 D HydroGel-coated slides from PerkinElmer in Boston, Massachusetts, provide a hydrophilic environment and the capacity to hold more than a single layer of probes.

To get you started, Genetix in New Milton, UK, has the QArraymini, a ready-to go benchtop microarray spotter. "People can just get it out of the box, and get on with it," says Gail Czerepinski, product manager for Genetix. The QArraymini is available as part of a protein-array package that includes source-plate chiller, buffers, epoxy-coated slides and scanner. The spotter uses contact pin printing to array antibodies or other proteins on 25 × 75 mm epoxy-coated slides, printing up to 120,000 spots on each slide. "People are printing between a few hundred and a few thousand protein spots. But they can easily scale up should they wish to," says Czerepinski. The aQuire confocal dual-laser scanner, with its own software, images Cy3 and Cy5 dyes simultaneously, with a blue laser fitted as an option.

Protein arrays made in individual wells of 96- or 384-well plastic microplates, such as the plates for the new high-throughput HTA platform from Greiner Bio-one in Frickhausen and Scienion in Berlin, Germany, are becoming popular, as they allow economical parallel processing of many different test samples. Genetix has developed an optional module for plate arraying for their high-throughput arrayers QArray2 and QArraymax.

L.M.



Ready-to-go: QArraymini gets you printing.

using any standard fluorescence scanner.

Fluorescence is not the only detection system on offer in kit arrays. The range of chemiluminescence-based TransSignal protein-domain arrays from Panomics in Redwood City, California, is expanding, with the addition of SH2 and WW domains. The FASTQuant kit from Schleicher and Schuell in Keene, New Hampshire, combines traditional enzyme-linked immunosorbent assay (ELISA) methodology with the power of multiplexing. Antibodies to nine different human cytokines are incorporated into an array on the company's FASTSlides — nitrocellulose-coated glass slides — and 64 arrays are arranged in a microtitre plate footprint. The company plans to release six FASTQuant kits for human and mouse cytokines, angiogenesis factors and human chemokines this year.

Adding content

One bottleneck that chip manufacturers face is content. "Getting an antibody against every protein in the human proteome is a gigantic task, more so than sequencing the genome. This is not something that will happen in the next decade," says Jan van Oostrum, head of proteome sciences in functional genomics at Novartis in Basel, Switzerland.

Increasing the content is a matter of competitive advantage for Molecular Staging in New Haven, Connecticut, which uses protein chips in their work with biotech and pharma companies to identify biomarkers for diseases and help reanalyse promising drugs that, although highly effective in some patients, failed clinical trials because of

unacceptable side-effects in others. "As soon as we see that a pathway is upregulated or downregulated we gain a mechanistic understanding of the disease, its progression, and patients' response to drugs," says Peter Fuller, senior vice-president of business development. "We work constantly to increase content — every four months we validate a new chip." The company's antibody-based chips use a signal-amplification technology developed at Yale University, New Haven. DNA circles attached to the detector proteins are replicated to produce a long DNA molecule that can be detected by attachment of large numbers of small detector molecules. Molecular Staging is collaborating with Eli Lilly to identify biomarkers for sepsis, a potentially deadly reaction to a blood infection that develops so rapidly that physicians find it difficult to diagnose quickly enough.

Multiplexing in solution

The challenges of constructing solid-surface arrays holding thousands of proteins with different properties are fuelling interest in protein-interaction assays in solution. Suspension-bead assays are particularly flexible, and can be adapted to both proteins and nucleic acids. The Bio-Plex system from Bio-Rad Laboratories in Hercules, California, uses Luminex's bead-based xMAP technology (see *Nature* 422, 917–923; 2003), as does the LiquiChip system from Qiagen Instruments in Hombrechtikon, Switzerland. For protein analysis, Bio-Plex uses antibodies from Cell Signaling Technologies

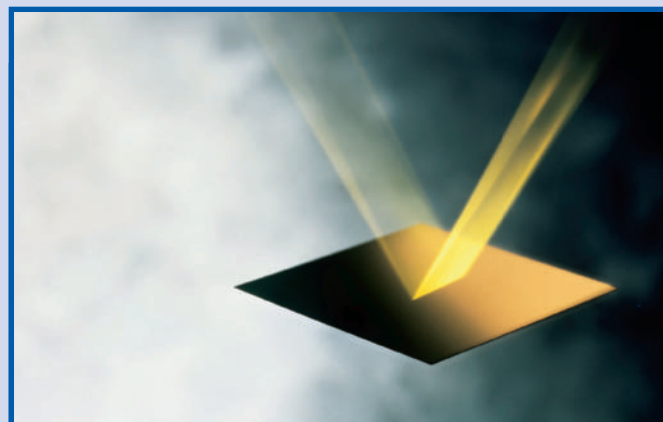
in Beverly, Massachusetts, to offer ready-to-go cytokine assays, and assays for phosphorylated signal transduction molecules, as well as customized arrays. Suspension-bead arrays are flexible enough to tackle any sort of protein–ligand interaction by simply coupling the required proteins or ligands to different bead populations. Luminex beads, for example, enable simultaneous quantitation of up to 100 different biomolecules in a single microplate well.

Another versatile platform is that of ACLARA Biosciences in Mountain View, California, which can support gene-expression and protein assays run side-by-side on the same sample. Detection is by proprietary eTAG reporter molecules, which are coupled to the ligands (for protein analysis these are usually antibodies), with a different eTAG for each ligand. Labelled ligands are applied to the sample in microtitre plates, or tissue sections on slides, and if the ligand binds to its target, the tag is released. Released tags are resolved, identified and quantified by capillary electrophoresis. An additional tacophoresis step concentrates the tags, taking detection sensitivity into the attomolar range. "It is possible to detect just 10 molecules," says Sharat Singh, chief technology officer at ACLARA. "In testing patient samples, we can predict whether a patient will respond or not by analysing a few analytes related to the mechanism of action of the drug and the biology of the specific target in the tumour."

The 3D HydroArray platform from Biocept in Carlsbad, California, uses a proprietary hydrogel polymer that is 95% water,

A NATURAL AFFINITY

How proteins interact with each other has become a key issue in drug discovery and development. Once a target for a drug is identified, it is also important to define how strongly it binds to other biomolecules to obtain information about selectivity. Biacore of Uppsala, Sweden, has been producing affinity-based sensor chips and instrumentation for protein-interaction analysis



Casting light on proteins: Biacore's gold-coated chip.

for more than a decade. The company's core technology is surface plasmon resonance (SPR), using a gold-coated glass chip onto which an array of protein-binding samples is immobilized. When the chip is illuminated, the interaction results in mass changes in the aqueous layer close to the sensor chip, that directly correlate to the refractive-index change. When molecules in the test solution bind to, or dissociate from, the immobilized protein, the refractive index rises and falls and a change is detected. Plotting the response against time allows continuous, real-time monitoring of interacting molecules. Because there is no need to label targets with fluorescent or radioactive labels, molecules can be studied in a near-native state, and the binding data closely reflect *in vivo* behaviour. Biacore has also developed the technology towards direct measurement of small-molecule binding to protein, for lead optimization in drug discovery, says Stephan Lofas, chief scientific officer. The company markets a variety of SPR instruments to suit different needs, from the high-throughput lab that requires unattended, fully automated runs for up to 384 samples at a time, to the smaller laboratory where researchers want to study a variety of different molecules.

The 8500 Affinity Chip Analyzer from Applied Biosystems, developed in conjunction with HTS Biosystems in Hopkinton, Massachusetts, is based on grating-coupled SPR technology. "If you are interested in target discovery, this system provides massively parallel detection of hundreds of binding events per analysis — thousands per day," says Enrico Picozza, chief technology officer at HTS Biosystems.

L.M.

enabling arrayed biomolecules (DNA or proteins) to keep their native conformation. Probe biomolecules are mixed with the pre-polymer, and nanolitre droplets are arrayed on a glass slide. During curing, covalent reactions simultaneously crosslink the pre-polymer, tether the biomolecules to the gel matrix and bind the droplets to the slide. The result is many levels of probes, giving high sensitivity and a broad dynamic range. Biocept has developed genomic arrays and protein arrays are coming soon.

Mapping interactions

“If you are talking proteomics, what you are looking for must not be predefined, it has to be an open approach,” says van Oostrum. Reverse arrays, where the cell extract itself is spotted on a chip and probed with a large number of antibodies, is such an approach. They allow researchers to track the activation or perturbation of cellular pathways, to monitor expression of disease-related proteins, and to investigate the cellular effects of drugs, all using very small amounts of sample. Zeptosens in Witterswil, Switzerland, for example, has developed the ZeptoMark CeLyA protein-profiling system, a sensitive reverse array that uses the company’s planar wave-guide technology. Each chip can contain the protein content of up to 384 different cell lysates for high-throughput profiling.

“We thought that after we had identified how many genes, how many proteins, the problem would go away,” says Celis. “But there are protein–protein interactions that give you different functions. Unless you



Zeptosens: microarrays must be made under rigorous clean-room conditions.

know those differences, it is difficult to come up with a drug therapy,” he points out.

Giulio Superti-Furga, vice-president of biology at drug-development company Cellzome in Heidelberg, Germany, says: “We look at proteins in their own juice, with their own partners, with their own post-translational modifications.” Cellzome uses two technologies based on liquid chromatography/tandem mass spectrometry to study the protein context of drugs and drug-like molecules. Potential drug compounds are immobilized and used as affinity reagents to identify binding proteins. Candidate drug targets are then ‘positioned’ onto cellular

pathways using tandem affinity purification, originally developed by researchers at the European Molecular Biology Laboratory in Heidelberg. “You can almost derive an organizational chart, as you’d draw for a company, but for a proteome, with its drug-binding components earmarked,” says Superti-Furga. Key to the approach is the compilation and visualization of data by informatics. Such maps should help identify points of intersection between different pathways, such as apoptosis and stroke for the TNF- α pathway and lipid metabolism for the APP pathway.

MDS Proteomics in Toronto, Canada, takes a chemiproteomics approach to iden-

FAMILY BUSINESS



ActivX: pulling out proteins by their activity.

“Suppose you could develop a chemical tool so powerful that you could go into a cell and identify all the members of one family — all the protein kinases, or all the hydrolases or the phosphatases,” proposes John Kozarich, president and chief scientific officer at the drug discovery and development company ActivX in La Jolla, California. The activity-based proteomics technology to do just that was originally developed by Ben Cravatt’s group at the Scripps Institute and has been commercialized by ActivX.

The approach is based on a simple idea: proteins in the same family tend to have conserved active sites which a well-designed chemical reagent can pick up. In the chemistry literature are thousands of publications probing the structure and functions of proteins, defining inhibitors and ligand interactions. “We harvest that information. We are interested in tools that allow us to look at families related by a common trait,” explains Kozarich.

For serine hydrolases, for example, ActivX has designed chemical tools that recognize the underlying catalytic motif — three amino acids in a particular spatial configuration. The system adds a highly fluorescent covalent tag to one of the unique amino-acid residues, and this allows all the serine hydrolases to be pulled out of a sample by capillary electrophoresis, and analysed and identified by mass spectrometry. Early in drug development, a domain scan can compare different animal models to help resolve toxicity issues. Another use is to characterize off-target interactions of drug leads, as the probes are designed to recognize just those protein motifs that are the same as the active sites that are the real drug target.

L.M.

tify primary and secondary protein targets of bioactive compounds, including those whose mechanism of action is unknown. Protein-drug complexes are pulled out of cells using a tag present on the drug. The proteins are then processed on a 'proteomics reactor', developed at the company, that simulates the physiological protein concentrations found in cells, making it more likely that the protein complexed with the drug can be chemically or biochemically processed and analysed by mass spectrometry.

A tough challenge for proteomics is the detection of integral membrane proteins which include important drug targets. By focusing on quantitative proteomic profiling of plasma-membrane proteins, drug-discovery company Caprion in Montreal, Canada, hopes to home in on the 1% of proteins in the cell that are likely to be the most relevant drug targets. Its proprietary Cell-Carta platform is an integrated and automated suite of technologies for fractionation of human tissue samples and plasma, quantitative analysis and directed mass spectrometry-based identification of protein targets.

Inside the intact cell

"The big issue is, what is physiological? What is going on *in vivo*? How do proteins come together and interact in intact cells?" says Sam Hanash of the University of Michigan Medical School, Ann Arbor, and president of the Human Proteome Organization. This international effort in proteomics focuses on organ systems and biological fluids relevant to diseases. "Some exciting technologies are

becoming available that may not be necessarily high-throughput but are more physiological." He cites methods that allow visualization of protein modifications in living cells, such as genetically encoded fluorescent indicators that detect protein-phosphorylation in signal transduction, for example.

A well-established technique for detecting protein-protein interactions is yeast two-hybrid analysis. "It's very powerful and you generate many interactions. But it is the first level of proteome mining; you then require more detailed follow-up and validation," says David Litman, senior vice-president of R&D and chief technology officer at BD Biosciences in San Jose, California, which makes the BD Matchmaker system for yeast two-hybrid analysis.

"A lot of proteomics is done in bulk, people taking cells and cracking them open to see what proteins are in there," says Litman. "But after scanning the whole proteome, one would want to analyse smaller sets of proteins." Antibodies to detect phosphorylated proteins are being commercialized by BD Biosciences as multiplexed immunoassays to interrogate intracellular pathways using flow cytometry. The new BD FACSArray bioanalyser accepts 96-well plates for high-throughput cellular analysis or multiplexed protein analysis using cytometric beads. A microtitre plate can be run in less than 35 minutes, detecting up to 1,000 events per well and reporting up to four



Sam Hanash:
keeping it real.

fluorescent and two optical scattering parameters. With this benchtop system, researchers can study protein interactions within a given pathway without having to separate cell populations before running the assay. Reagents to detect kinases, phosphorylated proteins and other activated protein states are available, and BD Biosciences is developing preconfigured kits.

Protein-capture microarrays are limited as they stand, partly because of nonspecific binding by the capture agents, says James Wang, chief technology officer at Hypermatrix in Worcester, Massachusetts. "Additional protein microarray formats are needed," he says. To increase specificity and expand array applications, Hypermatrix has developed the Staining AntibodyArray to simultaneously detect and localize large numbers of cellular antigens in intact cells. Detector antibodies are arrayed on one support, while the cells are attached to a second. When the two surfaces are apposed, antibodies dissociate from the array and bind to cell-surface antigens. After the supports are separated, secondary antibodies are used to profile the antigens revealed and their cellular location.

The proteomics boom continues. And although cutting-edge proteome analysis is expensive, the growing array of proteomics tools promises to supply researchers with a method suited to every experimental need. ■

Lisa Melton is science writer at the Novartis Foundation, London, UK.

BREAKING DOWN THE PROBLEM

A number of companies have developed peptide arrays for detecting protein interactions, such as the interaction of protein kinases with their substrates. Using high-throughput peptide synthesis, they manufacture an entire potential protein target as overlapping peptides and array them on a chip or a membrane. Pepscan Systems in Lelystad, the Netherlands, for example, offers the pre-arrayed PepChip, with 1,200 peptide substrates for protein kinase identification and assay, while New England Peptide in Gardner,

Massachusetts, has adopted the 96-well microplate format for its custom peptide arrays, with a different synthetic peptide in each well.

"The advantage is that smaller peptides are less dependent than proteins on secondary structure for their biological function," explains Holger Wenschuh of Jerini Peptide Technologies in Berlin, Germany. Jerini identifies peptide substrates for customers' orphan kinases, and for other categories of protein-modifying enzymes, such as proteases and phosphatases, using its PepStar microarray platform, in which up to 20,000 peptide sequences are synthesized on a membrane and then transferred to the complementary chip. The customer sends in their enzyme, Jerini systematically trawls the available literature and protein databases, such as Swiss-Prot and Phosphobase for potential substrates, sets up arrays, does the assay experiments and delivers the peptide substrate information, often in less than a week. Jerini also offers a ready-to-use PhosphoSite-Detector microarray kit to detect potential phosphorylation sites in kinase substrates. The readout is a phosphate-transfer event, which can be detected by autoradiography or by a phosphotyrosine-specific antibody.

Sigma's Michael Hadjisavas sees custom peptide arrays as "an area that is poised to explode". Custom peptide arrays using Sigma's PEPscreen platform are used to map epitopes, protein-protein interactions or protein-ligand interactions. "Since roughly half of all biotherapeutics are antibodies," says Hadjisavas, "it is important to demonstrate the specific peptide sequence that is being recognized in the target protein."

L.M.



Printing PepStar peptide chips.

JERINI PEPTIDE TECHNOLOGIES

Company	Products/activity	Location	URL
Protein and peptide capture, arrays, array equipment and multiplexing systems			
ACLARA BioSciences	eTAG detection technology for multiplexed protein and nucleic acid analysis	Mountain View, California	www.aclara.com
Advantix	ArrayBooster nanopump system for protein-chip and microarray incubation	Munich, Germany	www.advantix.de
Affibody	High-throughput protein identification and characterization by proprietary binding proteins	Stockholm, Sweden	www.affibody.com
Affymetrix	Spotter and arrayers	Santa Clara, California	www.affymetrix.com
Applied Precision	arrayWoRxe biochip reader	Issaquah, Washington	www.appliedprecision.com
Axon Instruments	GenePix 4000B microarray scanner and GenePix Pro microarray-analysis software	Foster City, California	www.axon.com
Biacore	Analysis and measurement of biomolecular interactions using surface plasmon resonance	Uppsala, Sweden	www.biacore.com
Bio-Rad	Instruments, software and consumables for life-sciences research; Bio-Plex bead-based multiplex analysis system for proteins and nucleic acids	Hercules, California	www.bio-rad.com
BMG Labtechnologies	Microplate and array readers and handling systems	Offenburg, Germany	www.bmg-labtechnologies.com
Caliper Life Sciences	LabChip microfluidic systems for drug discovery, biological and genetic research	Mountain View, California	www.calipertech.com
Cell Signaling Technology	Antibodies, including motif-specific antibodies	Beverly, Massachusetts	www.cellsignal.com
Ciphergen	ProteinChip system for protein identification and characterization from complex mixtures using SELDI technology	Fremont, California	www.ciphergen.com
Fluidigm	Microfluidics chip for protein crystallization	South San Francisco, California	www.fluidigm.com
Genetix	Picking, arraying, gridding, replicating, rearranging and microarraying robots	New Milton, UK	www.genetix.co.uk
Genomic Solutions	Automated instrumentation for genomics and proteomics, software, consumables and services; Investigator automated system for proteomics	Ann Arbor, Michigan	www.genomicsolutions.com
Greiner Bio-One	Consumables for high-throughput research and diagnostics, microplates for arrays	Frickenhausen, Germany	www.greinerbioone.com
Gyros	Gyrolab microfluidics CD for sample preparation for mass spectrometry, immunoassays	Uppsala, Sweden	www.gyros.com
HTS Biosystems	FLEX CHIP array platform for detection of protein interactions using grating-coupled surface plasmon resonance	Hopkinton, Massachusetts	www.htsbiosystems.com
Hypomatrix	Off-the-shelf AntibodyArrays and Staining AntibodyArrays for signal transduction proteins	Worcester, Massachusetts	www.hypomatrix.com
Jerini Array Technology	pepSTAR peptide chips for kinase and phosphatase target identification	Berlin, Germany	www.jerini.com
Molecular Staging	RCAT chip-based proteomics platform; whole-genome amplification kits	New Haven, Connecticut	www.molecularstaging.com
Nalge Nunc International	Labware, coated slides for DNA and protein microarrays	Rochester, New York	www.nalgenunc.com
NextGen Sciences	a2DE automated 2D gel electrophoresis workstation; ProteinArray workstation; expressionfactory automated gene-to-protein expression system	Huntingdon, UK	www.nextgensciences.com
Panomics	TranSignal Human Cytokine Antibody Array for 18 human cytokines.	Redwood City, California	www.panomics.com
Pepscan Systems	Peptide synthesis; epitope mapping; PepChip peptide chips for kinase targets	Lelystad, The Netherlands	www.pepscan.nl
PerkinElmer Life Sciences	Automated systems for liquid handling and sample preparation; slides for protein arraying; ProteinArray workstation	Boston, Massachusetts	lifesciences.perkinelmer.com
Protometrix	Yeast protein microarray for identification and interaction studies	Branford, Connecticut	www.protometrix.com
Scienion	sciCapture customized protein array; sciLISA antibody array; arrayer	Berlin, Germany	www.scienion.de
Sigma-Aldrich	Reagents and biochemicals for life sciences research; antibodies; Panorama Ab microarray for cell-signalling proteins	St Louis, Missouri	www.sigmaaldrich.com
Tecan	Laboratory automation for the life sciences, LabCD 'lab-on-a-disc' assays	Männedorf, Switzerland	www.tecan.com
TeleChem/ArrayIt International	ArrayIt products for microarray analysis; Protein Edition SpotBot arrayer, coated slides	Sunnyvale, California	www.arrayit.com
Zeptosens	SensiChip DNA microarray systems; ZeptoMARK protein-chip platform	Witterswil, Switzerland	www.zeptosens.com
Zymyx	Human cytokine profiling biochip system	Hayward, California	www.zymyx.com
General			
Advion BioSciences	Intellisense ESI chip for MS; contract liquid chromatography	Ithaca, New York	www.advion.com
Agilent	2100 Bioanalyzer workstation for RNA, DNA or protein analysis; instruments for genomics and proteomics	Palo Alto, California	www.agilent.com
Amersham Biosciences	Instruments, equipment and reagents for proteomics and cellular assays; Ettan range of equipment for sample preparation and mass spectrometry	Uppsala, Sweden	www5.amershambiosciences.com
Apogent Technologies	Labware and equipment for the life sciences	Portsmouth, New Hampshire	www.apogent.com
Applied Biosystems	DNA sequencers, mass spectrometers, workstations and automated instruments for genomics and proteomics	Foster City, California	home.appliedbiosystems.com
ARRM	Robotic spot cutter and plate replicator for proteomics	Adelaide, Australia	www.rrm.com
Beckman Coulter	Automated tools for molecular biology, biochemistry, genomics and proteomics; ProteomeLab	Fullerton, California	www.beckmancoulter.com
BD Biosciences	Culture media, radioassay kits, FACS range of flow cytometers; monoclonal antibodies, antibody arrays for proteomics; reagents and biochemicals for molecular biology	Franklin Lakes, New Jersey	www.bd.com
Biotrace International	Laboratory environmental monitoring equipment; culture media; instrumentation and reagents for microbiology	Bridgend, UK	www.biotrace.co.uk
Brinkmann	Laboratory instrument suppliers, software, consumables	Westbury, New York	www.brinkmann.com
Bruker Daltonics	Proteiner mass spectrometry-based proteomics suite	Billerica, Massachusetts	www.bdal.com
BSI Proteomics	Proprietary system for automated protein crystallization for crystallography	Gaithersburg, Maryland	www.bsiproteomics.com
Chromagen	Fluorescence-based products and services for genomics and proteomics	San Diego, California	www.chromagen.com
Cosmo Bio	Suppliers of monoclonal and polyclonal antibodies; analysis and measurement of biomolecules; culture media and equipment for biochemical research	Tokyo, Japan	www.cosmobio.co.jp
DataCentric Automation	Automated systems for protein crystallography	Nashville, Tennessee	www.dccorp.com
Dynal	Dynabeads for immunoprecipitation and phage-display panning	Oslo, Norway	www.dynalbiotech.com
Eppendorf	Laboratory instrumentation and consumables for molecular and cell biology	Hamburg, Germany	www.eppendorf.com
Finnzymes	Enzymes, PCR kits and transposon products for genomics and proteomics	Espoo, Finland	www.finnzymes.fi
Gilson	Automated equipment for genomics and proteomics	Middleton, Wisconsin	www.gilson.com

Company	Products/activity	Location	URL
Hamilton Company	Microlab automated liquid-handling workstations (Hamilton Bonaduz in Switzerland)	Reno, Nevada	hamiltoncomp.com
Invitrogen	Kits and reagents for cloning, genomics, proteomics, molecular and cell biology	Carlsbad, California	www.invitrogen.com
KBiosystems	Robotic instrumentation for high-throughput genomics and proteomics	Basildon, UK	www.kbiosystems.com
LI-COR Biosciences	Infrared imaging system for western blot and in-cell western blot assay analysis	Lincoln, Nebraska	www.licor.com
MDS Sciex	Mass spectrometer developers and producers	Concord, Ontario, Canada	www.sciex.com
Micromass-Waters	Micromass instruments for mass spectroscopy; HPLC systems; reagents and consumables	Milford, Massachusetts	www.waters.com
Millipore	Automated equipment for molecular biology, biochemistry, genomics and proteomics	Bedford, Massachusetts	www.millipore.com
Molecular Devices	Automated laboratory equipment	Sunnyvale, California	www.moleculardevices.com
Molecular Probes	Fluorescent probes	Eugene, Oregon	www.probes.com
MP Biomedicals	Reagents and biochemicals for life-sciences research	Aurora, Ohio	www.icnbiomed.com
Nanogen	NanoChip automated workstation for genomics and proteomics analysis	San Diego, California	www.nanogen.com
NextGen Sciences	a2DE automated 2D gel electrophoresis workstation	Huntingdon, UK	www.nextgensciences.com
Novagen	Reagents for genomics and proteomics	Madison, Wisconsin	www.novagen.com
Pierce	Components and consumables for automated systems in molecular biology, protein chemistry, sample handling, immunology, chromatography, and protein arrays	Rockford, Illinois	www.piercenet.com
Promega	Vectors, reagents and kits for genomics, proteomics and cellular analysis	Madison, Wisconsin	www.promega.com
Proteome Systems	Proteomics technology; ProteomeIQ integrated system for proteomics	North Ryde, Australia	www.proteomesystems.com
Qiagen	BioRobot multifunctional laboratory workstation and equipment for automated nucleic acid, protein and cell purification; reagents, kits and instruments for genomics and proteomics	Venlo, The Netherlands	www.qiagen.com
RoboDesign	Automated equipment for protein structure analysis	Carlsbad, California	www.robodesign.com
Roche Diagnostics	Reagents and kits for molecular biology, functional genomics and proteomics research	Lewes, UK	www.roche-applied-science.com
Shimadzu-Biotech	AXIMA mass spectrometers; XCISE integrated sample gel excision processor	Kyoto, Japan	www.shimadzu-biotech.net
Stratagene	Tools and reagents for genomics, proteomics, drug discovery and toxicology	La Jolla, California	www.stratagene.com
Thermo Finnigan	Instruments for chromatography and mass spectroscopy; ProteomeX workstation for automated LC/MS protein analysis	San Jose, California	www.thermo.com
Upstate	KinaseProfiler specificity testing, IC50 Profiler Express and services for crystallography assay development, bulk protein production and purification; immunoprecipitation kits	Charlottesville, Virginia	www.upstate.com

Contract services

Biocept	3D HydroArray hydrogel polymer spotted arrays	Carlsbad, California	www.biocept.com
Caprion Pharmaceuticals	Proteomics-based drug and biomarker discovery using proprietary CellCarta platform	Montreal, Canada	www.caprion.com
Cellzome	Proteomics-based target identification	Heidelberg, Germany	www.cellzome.com
SurroMed	Protein biomarker-based diagnosis and drug discovery	Mountain View, California	www.surromed.com
ActivX Biosciences	Proprietary chemical probes for protein families; drug discovery	La Jolla, California	www.activx.com
BioVision	Peptidomics: discovery and identification of peptides and small proteins	Hannover, Germany	www.biovision-discovery.de
Cytomx	Cell lines and proteins for signal transduction; molecular biology and proteomics services	Cambridge, UK	www.cytomx.com
Dualsystems Biotech	Custom yeast two-hybrid array services; yeast strains, plasmids, kits	Zurich, Switzerland	www.dualsystems.com
Eurogentec Proteomics	Validation of <i>in silico</i> and genomically derived targets	Berlin, Germany	www.eurogentec.com
GeneProt	Industrial-scale separation, identification and characterization of human proteins	Geneva, Switzerland	www.geneprot.com
Hybrigenics	<i>In silico</i> prevalidation, and experimental validation services	Paris, France	www.hybrigenics.com
MDS Proteomics	Target validation using proteomics-based systems biology	Toronto, Canada	www.mdsproteomics.com
Peptide Specialty Laboratories	Peptide synthesis	Heidelberg, Germany	www.peptide.de
Protagen	Protein analysis services; UNIClon human gene library	Dortmund, Germany	www.protagen.com
Proteome Sciences	Drug target and disease biomarker discovery	Cobham, UK	www.proteome.co.uk
Proteomic Solutions	Software for 2D gel analysis; distributor for Genomics Solutions; proteomics services	St Marcel, France	www.proteomicsolutions.fr
Trenzyme	Genomics and proteomics services	Konstanz, Germany	www.trenzyme.com

Databases and software

Accelrys	Discovery Studio products for database mining, genomics and proteomics	San Diego, California	www.accelrys.com
Array Genetics	Online bioinformatics protein information database and bioinformatics tools for proteomics	Newtown, Connecticut	www.arraygenetics.com
Confirmant	Protein Atlas of the Human Genome	Abingdon, UK	www.confirmant.com
Curagen	GeneScape portal for genome analysis tools; proteomics research for drug target discovery	Branford, Connecticut	www.curagen.com
DNASTAR	Desktop DNA and protein sequence analysis software	Madison, Wisconsin	www.dnastar.com
Genedata	Proprietary bioinformatics systems and services for sequence and genome analysis, functional genomics; database development and management	Basel, Switzerland	www.genedata.com
Incyte Genomics	Proteome BioKnowledge Library species-specific protein information databases	Palo Alto, California	www.incyte.com
Informax	Desktop bioinformatics software for Windows and Macintosh; contract services	Bethesda, Maryland	www.informaxinc.com
Inpharmatica	Biopendium and CeleraEdition Biopendium proteome annotation resources; PharmaCarta	London, UK	www.inpharmatica.com
LION Bioscience	Software for sequence and genome analysis, and proteomics, SRS-based DiscoveryCenter platform for integration of chemical and biological information	Heidelberg, Germany	www.lionbioscience.com
Matrix Science	MASCOT search engine for mass spectrometry data to identify proteins from sequence databases	London, UK	www.matrixscience.com
MegaMetrics	DIOGENES data-mining program for analysis of microarray, proteomics and SNP databases	Wyndmoor, Pennsylvania	www.megametrics.com
Nonlinear Dynamics	Phoretix image-analysis and database software for 1D and 2D gel electrophoresis and high-throughput arrays; Progenesis for the analysis of high-throughput 2D gels for proteomics	Newcastle-upon-Tyne, UK	www.nonlinear.com
PubGene	PubGene public access and commercial gene databases and analysis software for genomics and proteomics	Oslo, Norway	www.pubgene.com

● see advertisement

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequinot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejp.com

naturejobs

Swatting flies

It is five years since the National Academy of Sciences issued its seminal report on employment conditions for US postdocs. Last month, at a convocation in Washington, postdocs and policy-makers gathered to assess the progress being made over the issues identified by the academy's Committee on Science, Engineering, and Public Policy (COSEPUP), which produced the report.

Summing up the situation, Maxine Singer, chair of COSEPUP, noted that some postdoc programmes have improved pay, benefits, mentoring and career advice. But, she added, the underlying problems — notably a system that treats postdocs as cheap labour — have kept the majority from making significant gains.

Indeed, Singer compared efforts in the past five years to “swatting flies”, because the effects seem to be so diffuse and varied. For example, the stipend level at the US National Institutes of Health has risen, but the agency's National Research Service Award, widely viewed as a benchmark worldwide, hasn't always been followed (see *Nature* 427, 178–179; 2004). And benefits such as health insurance and child care vary widely from institution to institution, even though most of the funding for postdocs comes from a few common sources.

“Maybe we should look at fundamental issues,” Singer suggested. Her thoughts on reform include setting strict deadlines for how long a postdoc can last, limiting the number of fellows in each lab and increasing the number of postdocs funded through independent grants rather than through a principal investigator.

These changes will make some senior scientists uncomfortable because they would have to relinquish some of their power. But ceding some control to postdocs would ultimately benefit both parties because more young scientists would be more willing to stay on an academic career path.

Paul Smaglik
Naturejobs editor



Contents

CAREER VIEW

Nuts & Bolts

Passing the interview

Graduate Journal

Through the looking glass

Movers

Alistair Walker

p112

WWW.NATUREJOBS.COM

Career centre

Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

GRADUATE JOURNAL

Through the looking glass

The English countryside flew by as I sat on my train heading for Cambridge, the university where I spent four years as an undergraduate. This time I was going there as a postgrad to take part in a multidisciplinary retreat.

Once there, I found myself with 80 other PhD students. We were an eclectic bunch: scientists, historians, theologians, psychologists, mathematicians and composers. Over a few days, we had to team up to manage an environmental crisis, help a developing nation, and even run the European Commission.

When you throw strangers together and attempt to solve such problems, some amazing things happen. People fell into different roles quite naturally — leaders, thinkers, coordinators — and I found the team dynamics fascinating to watch. Many students work in relative isolation, so working intensively with others for 12 hours a day was a real eye-opener.

An odd thing for a PhD student to do, you might say, but it was a wonderful experience. I gained insight into myself and others, and made new friends. By the end of the week I was shattered but I didn't want to leave. As my train pulled away, the countryside was the same but I was looking at it through slightly different eyes. ■

Amber Jenkins is a graduate student in particle physics at Imperial College, London, doing thesis research at Fermilab in Batavia, Illinois.

NUTS & BOLTS

Passing the interview

So, you've done your homework on the organization, applied for a job and landed an interview. How can you give yourself the best possible chance of success? The key is preparation.

Before you go, pick three to five points that you believe are essential for you to make in the interview. This will allow you to highlight these messages if you're asked open-ended questions such as: "Tell me about yourself" or "Why should we hire you?"

You also need to be ready to deal with 'behaviour-based' questions. These will focus on your past behaviour in an effort to predict your future performance. They are designed to make you give specific examples of the competencies and behaviours that the interviewer is seeking. To respond well, you should cover three aspects:



With Deb Koen
Careers consultant

situation, action and result. For example, rather than simply stating that you're resourceful, offer a specific example that shows this.

Early on in the interview, you should identify the interviewer's main priorities. Assess the most important issues and address them in your responses. Capturing the interviewer's perspective will allow you to present your strengths in a way that matches their needs.

Don't talk for the sake of talking. Spend at least half of your time listening so that you better understand the requirements and

underlying needs of the position. Asking questions about the corporate culture and development opportunities will allow you to assess your compatibility for the post.

Once the interview is over, remember to follow it up. If you are still interested in the job, close the interview by expressing your enthusiasm. Find out the timeline for the appointment, and encourage the interviewer to contact your references. Send a letter reinforcing your interest. About a week later, call the company to check on the vacancy's status and again express your interest.

A strong presentation coupled with questions that reflect interest and insight will put you in a good position to get an offer and make a decision that is best for you. ■

Deb Koen is vice-president of Career Development Services and a columnist for The Wall Street Journal's CareerJournal.com.

MOVERS

Alistair Walker, director, Cerro Tololo Inter-American Observatory, La Serena, Chile



To follow a waxing interest in astronomy, Alistair Walker had to leave his native New Zealand. He chose South Africa as a place to do his PhD, partly because the University of Cape Town's astronomy department was one of only a few places that used the computer-controlled instrumentation that had captured Walker's imagination.

As an undergraduate, Walker had heard of high-speed photoelectric photometry, a new technique for rapidly measuring light emitted by stars and storing the data digitally. This technique was a revolution in the early 1970s, when

basic photographic work was often the norm in astronomy, says Walker.

The technology was developed by Edward Nather at the University of Texas and applied to astronomy by Brian Warner. Walker found it "technically exciting", so when he heard that Warner had moved to Cape Town, he jumped at the chance to work with the pioneer.

That turned out to be a good decision because, in addition to tapping into Warner's expertise, the Astronomical Observatory's Sutherland facility provided him with ample observing time and state-of-the-art equipment. But the distance and isolation proved trying at times. He travelled from New Zealand to "spectacularly beautiful" Cape Town by ship, and never went to a single international conference during his PhD. "That would be very unusual these days," Walker says.

After that, he bounced back and forth between South Africa and South America

— often drawn, as before, by new technologies and techniques. For example, he returned to South Africa following the development by Alec Boksenberg of University College London of an image photon counting system (IPCS) to improve two-dimensional spectroscopy. Boksenberg had built two machines, one of which was designated for the William Herschel Telescope in La Palma, Spain. But because that telescope was then two years from operation, the IPCS was sent instead to the South African Astronomical Observatory, so Walker followed.

Walker says that it is important to identify technology in its infancy and "to take a chance". He has even passed up an opportunity to work closer to home, in Australia, choosing Chile instead, because he thought he had better scientific opportunities there. And he is happy to have built his entire scientific career south of the Equator. ■

CV **1987–:** Cerro Tololo Inter-American Observatory (CTIO) in Chile — member of scientific staff since 1987, deputy director since 2000
1979–87: staff scientist, South African Astronomical Observatory, Cape Town
1977–79: Assistant scientist, CTIO